The Journal of Abnormal
and
Social Psychology

Editor: M. BREWSTER SMITH, *New York University*
Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University
URIE BRONFENBRENNER
Cornell University
ROGER BROWN
Massachusetts Institute of Technology
DONALD T. CAMPBELL
Northwestern University
LAUNOR F. CARTER
System Development Corporation
RICHARD CHRISTIE
Columbia University
MORTON DEUTSCH
Bell Telephone Laboratories
I. E. FARBER
State University of Iowa
LEON FESTINGER
Stanford University
DONALD W. FISKE
University of Chicago
ROBERT R. HOLT
New York University
IRVING L. JANIS
Yale University
HAROLD E. JONES
University of California
HAROLD H. KELLEY
University of Minnesota
LEONARD S. KOGAN
Community Service Society of New York
DANIEL J. LEVINSON
Harvard Medical School
GARDNER LINDZEY
University of Minnesota
DAVID C. MCCLELLAND
Harvard University
LEO POSTMAN
University of California
ELIOT H. RODNICK
Duke University
NEVITT SANFORD
University of California
SEYMOUR SARASON
Yale University
RICHARD L. SOLOMON
Harvard University
RENATO TAGIURI
Harvard University

Managing Editor: ARTHUR C. HOFFMAN
Promotion Manager: HELEN ORR
Editorial Assistant: BARBARA CUMMINGS

VOLUME 57
1958

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.
MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.
AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6, D.C.

Copyright 1958 by the American Psychological Association, Inc.

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated 1.8.70
Acc. No. 8280

CONTENTS OF VOLUME 57

ARTICLES, CASE REPORTS, CRITIQUE AND NOTES

PAGE

Adelson, J., & Redmond, Joan. Personality Differences in the Capacity for Verbal Recall 244
Anonymous. A New Theory of Schizophrenia 226
Baer, D. M. See Gewirtz, J. L.
Bender, I. E. Changes in Religious Interest: A Retest After 15 Years 41
Binder, A. Personality Variables and Recognition Response Level 136
Blake, R. R. See Coleman, Janet F.
Boardman, W. K. See Weinstein, A. D.
Bradford, N. See Gollin, E. S.
Budin, W. See Flavell, J. H.
Burdick, H. A., & Burnes, A. J. A Test of "Strain Toward Symmetry" Theories 367
Burnes, A. J. See Burdick, H. A.
Buss, A. H., & Gerjuoy, Irma R. Verbal Conditioning and Anxiety 249
Carson, R. C. Intralist Similarity and Verbal Rote Learning Performance of Schizophrenic and Cortically Damaged Patients 99
Cavanaugh, D. K. Improvement in the Performance of Schizophrenics on Concept Formation Tasks as a Function of Motivational Change 8
Clark, A. See Wallin, P.
Coleman, Janet F., Blake, R. R., & Mouton, Jane S. Task Difficulty and Conformity Pressures 120
Conger, J. J., Sawrey, W. L., & Turrell, E. S. The Role of Social Experience in the Production of Gastric Ulcers in Hooded Rats Placed in a Conflict Situation 214
Davids, A., & Parenti, Anita N. Time Orientation and Interpersonal Relations of Emotionally Disturbed and Normal Children 299
Davidson, K. S. See Waite, R. R.
Danziger, K. Value Differences Among South African Students 339
deCharms, R. See Jones, E. E.
Dixon, N. F. The Effect of Subliminal Stimulation Upon Autonomic and Verbal Behavior 29
Draguns, J. See Flavell, J. H.
Duncan, C. P. See Lewis, D. J.
Eysenck, H. J. Hysterics and Dysthymics as Criterion Groups in the Study of Introversion-Extraversion: A Reply 250
Fager, R. E., & Knopf, I. J. Relationship of Manifest Anxiety to Stimulus Generalization 125
Feinberg, L. D. See Flavell, J. H.
Fink, D. R., Jr. Negative Evidence Concerning the Generality of Rigidity 252
Fisher, S. Body Image and Asymmetry of Body Reactivity 292
Flavell, J. H., Draguns, J., Feinberg, L. D., & Budin, W. A Microgenetic Approach to Word Association 1
Franks, C. M. See Sigal, J. J.
French, Elizabeth G. The Interaction of Achievement Motivation and Ability in Problem-Solving Success 306
Frijda, N. H. Facial Expression and Situational Cues 149
Gardner, C. W. See Rosenberg, M. J.
Gerard, H. B. Some Effects of Involvement Upon Evaluation 118
Gerjuoy, Irma R. See Buss, A. H.
Gewirtz, J. L., & Baer, D. M. Deprivation and Satiation of Social Reinforcers as Drive Conditions 165
Ghiselli, E. E., & Lodahl, T. M. Patterns of Managerial Traits and Group Effectiveness 61
Goldstone, S. See Weinstein, A. D.
Gollin, E. S., & Bradford, N. "Faulty" Communication and the Spiral Aftereffect: A Methodological Critique 122
Gordon, J. E. See Hokanson, J. E.
Gourevitch, Susannah. See Klein, G. S.
Hall, R. L. See Rosenberg, S.
Heilbrun, A. B., Jr. Vocabulary Response as Related to Lateralization of Cerebral Lesion: An Investigation of "Latent Aphasia" 237
Hemphill, J. K. See Pepinsky, Pauline N.
Hokanson, J. E., & Gordon, J. E. The Expression and Inhibition of Hostility in Imaginative and Overt Behavior 327
Holt, R. R. See Klein, G. S.
Jackson, J. M., & Saltzstein, H. D. The Effect of Person-Group Relationships on Conformity Processes 17
Jenkin, Noel. Size Constancy as a Function of Personal Adjustment and Disposition 334
Jones, E. E., & deCharms, R. The Organizing Function of Interaction Roles in Person Perception 155
Jones, E. E., & Kohler, Rika. The Effects of Plausibility on the Learning of Controversial Statements 315

PAGE

Jones, E. E., Wells, H. H., & Torrey, R. Some Effects of Feedback From the Experimenter on Conformity Behavior 207

Kalnins, D. See Lindzey, G.

Karon, B. P. See Rosberg, J.

Kipnis, D. The Effects of Leadership Style and Leadership Power Upon the Inducement of an Attitude Change 173

Klein, G. S., Spence, D. P., Holt, R. R., & Gourevitch, Susannah. Cognition Without Awareness: Subliminal Influences Upon Conscious Thought 255

Knopf, I. J. See Fager, R. E.

Kohler, Rika. See Jones, E. E.

Lewis, D. J., & Duncan, C. P. Vicarious Experience and Partial Reinforcement 321

Lighthall, F. F. See White, R. R.

Lindsley, O. R. See Mednick, Martha T.

Lindzey, G., & Kalnins, D. Thematic Apperception Test: Some Evidence Bearing on the "Hero Assumption" 76

Lindzey, G., Tejessy, Charlotte, & Zamansky, H. S. Thematic Apperception Test: An Empirical Examination of Some Indices of Homosexuality 67

Lodahl, T. M. See Ghiselli, E. E.

Long, J. See McNamara, H. J.

Lovaas, O. I. Social Desirability Ratings of Personality Variables by Norwegian and American College Students 124

McNamara, H. J., Solley, C. M., & Long, J. The Effects of Punishment (Electric Shock) on Perceptual Learning 91

Mark, H. J., & Pasamanick, B. Variability of Light Perception Thresholds in Brain-Injured Children. 25

Mathews, Anne, & Wertheimer, M. A "Pure" Measure of Perceptual Defense Uncontaminated by Response Suppression 373

Mednick, Martha T., & Lindsley, O. R. Some Clinical Correlates of Operant Behavior 13

Morrison, H. W. See Peak, Helen.

Mouton, Jane S. See Coleman, Janet F.

Olson, G. W. Failure and the Subsequent Performance of Schizophrenics 310

Osgood, C. E. See Triandis, H. C.

Parenti, Anita N. See Davids, A.

Pasamanick, B. See Mark, H. J.

Peak, Helen, & Morrison, H. W. The Acceptance of Information Into Attitude Structure 127

Peck, R. F. Family Patterns Correlated With Adolescent Personality Structure 347

Pepinsky, Pauline N., Hemphill, J. K., & Shevitz, R. N. Attempts to Lead, Group Productivity, and Morale Under Conditions of Acceptance and Rejection 47

Phillips, L., & Rabinovitch, M. S. Social Role and Patterns of Symptomatic Behaviors 181

Pisoni, Stephanie. See Salzinger, K.

Purcell, K. Some Shortcomings in Projective Test Validation 115

Rabinovitch, M. S. See Phillips, L.

Rechtschaffen, A. Neural Satiation, Reactive Inhibition, and Introversion–Extraversion 283

Redmond, Joan. See Adelson, J.

Rosberg, J., & Karon, B. P. The Oedipus Complex in a Case of Deteriorated Schizophrenia 221

Rosenberg, M. J., & Gardner, C. W. Some Dynamic Aspects of Posthypnotic Compliance 351

Rosenberg, S., & Hall, R. L. The Effects of Different Social Feedback Conditions Upon Performance in Dyadic Teams 271

Saltzstein, H. D. See Jackson, J. M.

Salzinger, K., & Pisoni, Stephanie. Reinforcement of Affect Response of Schizophrenics During the Clinical Interview 84

Sarason, S. B. See Waite, R. R.

Sawrey, W. L. See Conger, J. J.

Shevitz, R. N. See Pepinsky, Pauline N.

Sigal, J. J., Star, K. H., & Franks, C. M. Hysterics and Dysthymics as Criterion Groups in the Measure of Introversion–Extraversion: A Rejoinder to Eysenck's Reply 381

Sigal, J. J., Star, K. H., & Franks, C. M. Hysterics and Dysthymics as Criterion Groups in the Study of Introversion–Extraversion 143

Smith, A. J. Perceived Similarity and the Projection of Similarity: The Influence of Valence 376

Smith, H. T. A Comparison of Interview and Observation Measures of Mother Behavior 278

Solley, C. M. See McNamara, H. J.

Spence, D. P. See Klein, G. S.

Staats, A. W., & Staats, Carolyn K. Attitudes Established by Classical Conditioning 37

Staats, Carolyn K. See Staats, A. W.

Star, K. H. See Sigal, J. J.

PAGE

TAYLOR, JANET A. The Effects of Anxiety Level and Psychological Stress on Verbal Learning...... 55
TEAHAN, J. E. Future Time Perspective, Optimism, and Academic Achievement.................. 379
TEJESSY, CHARLOTTE. See LINDZEY, G.
TORREY, R. See JONES, E. E.
TRIANDIS, H. C., & OSGOOD, C. E. A Comparative Factorial Analysis of Semantic Structures in Monolingual Greek and American College Students.................................... 187
TURNER, R. H., & VANDERLIPPE, R. H. Self-Ideal Congruence as an Index of Adjustment............ 202
TURRELL, E. S. See CONGER, J. J.
VANDERLIPPE, R. H. See TURNER, R. H.
WAHLER, H. J. Response Selections of Neurotic and Control Subjects as a Function of Relative Familiarity and Social Desirability of Items.................................... 239
WAITE, R. R., SARASON, S. B., LIGHTHALL, F. F., & DAVIDSON, K. S. A Study of Anxiety and Learning in Children.................................... 267
WALLIN, P., & CLARK, A. Marital Satisfaction and Husbands' and Wives' Perception of Similarity in Their Preferred Frequency of Coitus.................................... 370
WALTERS, R. H. Conditioning of Attention as a Source of Autistic Effects in Perception............ 197
WALTERS, R. H. The Intelligence Test Performance of Maori Children: A Cross-Cultural Study.... 107
WEINSTEIN, A. D., GOLDSTONE, S., & BOARDMAN, W. K. The Effect of Recent and Remote Frames of Reference on Temporal Judgments of Schizophrenic Patients.................................... 241
WELLS, H. H. See JONES, E. E.
WERTHEIMER, M. See MATHEWS, ANNE.
ZAMANSKY, H. S. See LINDZEY, G.

A MICROGENETIC APPROACH TO WORD ASSOCIATION[1]

JOHN H. FLAVELL, JURIS DRAGUNS, LEONARD D. FEINBERG, AND WILLIAM BUDIN
University of Rochester

In a recent paper, Flavell and Draguns (2) have sketched the history of the concept of microgenesis—the series of events presumed to occur in the course of a single, brief conceptual or perceptual act—and have attempted to derive, as first steps in the formulation of a microdevelopmental approach, certain tentative propositions about such transient evolutions. In the particular case of word association, theoretical considerations suggest that there are covert word responses that push for expression early in the associative process and that these early microgenetic forms are less logical, more "paleological" (1) in character than later ones. It is further assumed that when normal individuals associate in a relatively unhurried fashion, these "immature" responses are usually suppressed, the microgenetically later, more logical associations being the ones actually spoken. If these assumptions are correct, the likelihood of such immature responses being given overtly by normal individuals should increase under conditions that force very rapid response. Another current working hypothesis of the microgenetic approach, following Schilder (7), suggests that schizophrenic thinking can be characterized as microgenetically immature cognition. That is, schizophrenics tend to give overt expression as a final product to early cognitions that normals usually suppress in favor of the later, more mature cognitive responses.

It is the primary purpose of the study to make a preliminary test of two broad hypotheses derived from these speculations. First, normal subjects instructed to associate very quickly should give more microgenetically immature responses and fewer mature ones in word association than should comparable normal subjects performing without extreme time pressure. Second, schizophrenics should likewise give more immature and fewer mature associations than comparable normals, under conditions that do not involve pressure to respond immediately. The first two experiments to be described constitute attempts to explore the first hypothesis; the third experiment bears upon the second.

EXPERIMENT I

Method

Eighty-four university undergraduates served as *S*s: 52 women and 32 men. The apparatus consisted of a .01 sec. Standard Electric Timer, started and stopped by a hand key, and a Selmer Metronoma electric metronome.

All 84 *S*s were first given a 61-item word association test and all protocols were scored in accordance with the scoring system to be described in the next section.[2] The *S*s were then divided into three groups of 28 each —*Slow*, *Fast*, and *Fast-Distraction*—roughly equated for sex and for mean number of responses scored "microgenetically immature" on this pretest word list.

A second association test of 79 common words was then administered individually to each of the 84 *S*s. The first three authors served as *E*s, each taking a roughly equal share in testing of *S*s of each sex within each group. The *S*s were blindfolded throughout the test in order to avoid distraction. *E* read aloud instructions and stimulus words and recorded reaction times to the nearest .01 sec. by means of the hand key and electric timer. The key was depressed when *E* began to say the stimulus word and was released when *S* started to say the response word. *Slow* *S*s were instructed to respond to each stimulus word with the first word which came to mind. However, they were also told not to rush or hurry the association process; rather, they were instructed to sit back, relax, and wait for the first association to come to mind. Reaction times were to be taken, they were told, but only "for the record." The instructions for the *Fast* group, on the other hand, stressed repeatedly the paramount importance of responding as quickly as possible. These *S*s were told that, above all, they were to respond with a word just as quickly as they possibly could. Members of the third, *Fast-Distraction* group[3] were given speed-oriented instructions as were the *Fast* *S*s but, in addition, had

---

[1] The authors wish to thank Eugene Sachs for his assistance in collecting a portion of the data.

[2] Word association lists, instructions to *S*s, and full scoring instructions for association categories have been deposited with the American Documentation Institute. Order Document No. 5461, remitting $1.25 for 35-mm. microfilm or $1.25 for 6 by 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[3] The purpose of including the *Fast-Distraction* group in the study was twofold. First, it constituted an additional "time-pressure" group with which to compare the *Slow* group. Second, it provided an opportunity to find out whether time pressure coupled with distraction yielded more immature responses than time pressure alone (*Fast* group).

also the distracting task of making a pencil stroke at every third beat of a metronome set to click twice per second throughout the administration of the test. Both *Fast* and *Fast-Distraction* Ss were from time to time urged by *E* to increase their speed of response in the course of presenting the first 13 stimulus words. These 13 words were thus treated as "practice words" and only the responses to the remaining 66 words were used in the statistical analyses.

Scoring

A number of investigators have described scoring systems designed to assess the formal relationships between stimulus and response words (3, 4, 6, 9). From these systems, the writers selected 16 categories which, with modifications, seemed to permit relatively easy and unambiguous scoring. These word association categories were classified on theoretical grounds into three types: first, those presumed to reflect microgenetically early or immature cognition; second, those presumed to reflect cognition in the later, final phases of microdevelopment; third, certain miscellaneous categories of indeterminate microgenetic phase that had been used in previous word association studies and appeared to deserve further empirical study. A brief description of the 16 categories follows.

A. *Immature Categories:*

1. *Completion.* The response word completes a word or common phrase of which the stimulus word is a part (needle—"haystack," glow–"worm").

2. *Distant.* The response word is either unrelated or distantly related in meaning to the stimulus word, or else is a decidedly idiosyncratic, personal association which may possess a moderate semantic relationship to the stimulus word (love—"Mary," book—"darkness").

3. *Perseveration.* The response word is unrelated or only distantly related to its stimulus word but is either obviously related to the stimulus or response word just preceding it in the series or else is identical to some stimulus or response word given earlier in the series.

4. *Perseveration-meaningful.* The response word is identical to some stimulus or response word given earlier but is also closely related to its stimulus word.

5. *Clang.* The response word seems to have been primarily achieved by means of a process of association based upon physical rather than semantic relationships (cold—"gold," glow—"fast"[4]).

6. *Emotional.* The response word clearly connotes an affective, evaluative judgment by *S* with respect to the referent of the stimulus word (love—"good," snake—"ugh!')

7. *Repetition.* The response word is a repetition or a partial repetition of the stimulus word, or is a grammatical variant of the stimulus word (expectation—"expect," deep—"depth").

8. *Multi-word-discrete.* The *S* gives two or more response words to a single stimulus word and these response words denote different or discrete concepts rather than constituting a multi-word expression of a single concept (jewel—"ruby," then "diamond").

B. *Mature Categories:*

1. *Synonym.* The response word is an almost exact synonym of the stimulus word (liberty—"freedom").

2. *Supraordinate.* The response word denotes a class of which the stimulus word's referent is a class member (egg—"food").

3. *Subordinate.* The response word denotes a member of the class signified by the stimulus word (fruit—"apple").

C. *Indeterminate Categories:*

1. *Attribute.* The response word is an adjective and the stimulus word a noun or vice-versa, providing only that the adjective could actually modify the noun in question (deep—"ocean," but not, bed—"sleepy").

2. *Verb.* The response word is a verb and the stimulus word a noun or vice-versa, providing only that the action denoted by the verb could actually be carried out by, or upon, the object signified by the noun (snore—"father," cheese—"eat").

3. *Contrast-coordinate.* The response word is either an antonym of the stimulus word or a coordinate, complimentary term on about the same level of abstraction (dark—"light," salt—"pepper," wall—"ceiling").

4. *Multi-word.* The *S* gives two or more response words to a single stimulus word but these response words all express a single concept (citizen—"a person who pays taxes").

5. *Blocking.* A response is either not offered at all or else has a latency greater than 10 seconds.

Rationale

The rationale for assigning particular categories to early versus late microdevelopmental phases was briefly this. *Synonym*, *Supraordinate*, and *Subordinate* are thought to reflect late phases because of their obvious logical, "secondary process" character. *Clang* and *Repetition* responses are believed to give evidence of a very beginning associative sequence, one which has not progressed beyond mere apprehension of the physical characteristics of the stimulus word. *Perseveration*, *Perseveration-meaningful*, and *Completion* suggest a contextual rather than strictly logical connection between the stimulus word and either another word with which it is frequently and mechanically paired in everyday life (*Completion*) or some word still echoing from a previous association (*Perseveration* and *Perseveration-meaningful*). Similarly, it is early rather than late in thought development that mediated and affect driven associations occur (*Distant*), that a number of discrete associative directions may exist without resolution into a single direction (*Multi-word-discrete*), and that primitive, dichotomous judgments of a "like—dislike" or "good—bad" character prevail (*Emotional*).[5]

---

[4] When *S* has given "fast" in response to "slow" just previously, the presumption being that "glow" has here been assimilated to "go."

[5] For a more detailed account of the assumed characteristics of early versus late developmental phases, see Flavell and Draguns (2). It might be mentioned here, however, that the present assignment of response categories to one or another phase stems from the authors' own (and perhaps idiosyncratic) integration of a variety of existing statements about the microdevelopmental process. Hence, the present interpretation of any given category may differ in some particulars from any single previous interpretation of similar categories. As an example, the *Distant* category is both defined and interpreted somewhat differently from the response class of the same name established by Rapaport, Gill, and Schafer (6).

Scoring Conventions

In using the scoring system, several conventions were adopted. Following Rapaport, Gill, and Schafer (6), the system permits multiple scoring of a single response. Two limitations were imposed, however. First, if a response could be scored for both an immature and a mature or indeterminate category, it was scored for the immature category only. For example, if a response word were a synonym for a stimulus word but had also occurred earlier in the series, it was scored *Perseveration-meaningful*, not *Synonym*. Also, a response was not scored for both of two categories if the two categories were not independent by definition. For instance, a response scored *Perseveration* would by definition also be a *Distant* response and would therefore be scored only for the more specific category, that is, *Perseveration*. It was found that about 80 per cent of all responses given received at least one scoring in this system. It should be emphasized that despite strenuous attempts to rid it of scoring ambiguities, the present scoring system still possesses numerous imperfections. Thus, it was sometimes difficult to decide between *Distant* and not *Distant*, *Perseveration* and *Perseveration-meaningful*, *Completion* and *Contrast-coordinate*, and *Emotional* and *Attribute*, to take but a few examples. In order to get some estimate of the interjudge reliability of the scoring system, 11 randomly chosen protocols were independently scored by two of the *E*s. There was agreement on 589, or 80 per cent of the 733 scorings made. This level of agreement, although not as high as would be desirable, compares favorably with the average agreement of about 67 per cent achieved by Karwoski and Berthold (4) using half as many scoring categories. Recognizing that interpretative biases of the scorer must inevitably enter into the scoring, it was decided that only one *E* (the senior author) should score all protocols in order to make the bias more or less constant rather than variable across groups.[6] Further, each *E* conducted a brief inquiry, following administration of the test, on those responses which he felt might raise difficult scoring problems later.

TABLE 1

WORD ASSOCIATION RESPONSE LATENCY FOR *SLOW*, *FAST*, AND *FAST-DISTRACTION* GROUPS
(Time in Seconds)

Measure	*Slow*	*Fast*	*Fast-Distraction*
Mean	2.27	1.60	1.55
SD	0.66	0.23	0.26

TABLE 2

CATEGORY SCORES FOR *SLOW*, *FAST*, AND *FAST-DISTRACTION* GROUPS

Category	*Slow*		*Fast*		*Fast-Distraction*	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Immature						
Completion	4.75	4.17	5.86	3.79	5.60	1.30
Distant	2.36	1.85	3.04	2.07	3.39	2.18
Perservation	1.46	1.37	4.21	4.47	3.57	3.67
Perseveration-meaningful	6.39	3.49	7.64	3.64	7.79	3.37
Clang	0.43	0.70	1.10	1.59	0.93	0.99
Emotional	0.75	1.49	1.39	1.95	0.61	1.10
Multi-word-discrete	1.68	3.66	1.21	1.47	0.96	1.42
Repetition	0.75	1.22	0.57	0.80	0.82	0.96
Mature						
Synonym	5.04	2.44	4.50	2.25	4.64	1.89
Supraordinate	0.43	0.58	0.89	1.08	0.57	0.70
Subordinate	2.00	1.47	1.21	1.10	1.18	1.04
Indeterminate						
Attribute	5.86	3.62	4.25	2.61	3.71	2.30
Verb	3.82	2.49	2.54	1.50	3.54	2.59
Contrast-coordinate	12.60	5.67	14.93	5.18	14.64	4.30
Multi-word	1.64	2.34	0.93	0.96	1.54	3.07
Blocking	0.39	0.92	0.57	1.24	0.79	1.10

Results

Table 1 shows the reaction time data for the three groups. As was expected, both *Fast* and *Fast-Distraction* groups showed significantly briefer mean latencies ($P < .01$) than the *Slow* group (t = 5.04 and 5.38, respectively).[7] The differences in mean latency between *Fast* and *Fast-Distraction* groups were not statistically significant.

The actual word association response data were analyzed by means of a three-way chi-square analysis of variance recently described by Wilson (10). Separate analyses on each of the 16 categories were performed for group (three) times sex (two) times experimenter (three). None of the main effects for sex and experimenter proved statistically significant, although there were significant but uninterpretable sex-experimenter interactions ($P < .05$) for two categories—*Completion* and *Repetition*. With regard to main effects for group, it was predicted that the *Fast* and *Fast-Distraction* groups would exceed the *Slow* group in frequency of immature responses and the converse for mature responses. Only two main effects for group were actually found to be statistically reliable: *Perseveration* ($P < .001$) and *Distant* ($P < .01$). Individual chi-square analyses for these two categories yielded the following findings: first, for *Perseveration*, *Fast* greater than *Slow* ($P < .001$) and *Fast-Distraction* greater than *Slow* ($P < .10$); second, for *Distant*, *Fast* greater than *Slow* ($P < .01$) and *Fast-Distraction* greater than *Slow* ($P < .01$). Table 2 shows Means and *SD*s, by category, for *Slow*, *Fast* and *Fast-Distraction* groups.

---

[6] It is recognized that this procedure does not in itself guard against the kind of bias effects wherein the scorer, being at the same time an *E*, may unconsciously resolve difficult scoring problems in favor of his hypotheses in cases where the group identity of the protocol is known to him—or, conversely, "lean over backwards" to err systematically against his hypotheses. In the present study, the scorer—as he scored—unfortunately knew the group identity of all the protocols in Experiment II and a few of the protocols in Experiments I and III. To be sure, certain precautions were taken. In cases where the group identity was known, the scorer resolved really borderline scoring problems conservatively, i.e., against his hypothesis; in Experiment II the scorer compared each *S*'s *Fast-Reward* protocol to his previous *Slow* one and automatically scored identical responses in each protocol identically; and throughout, the scorer continuously checked back to previous protocols to make sure his scoring standards remained as constant as possible. These precautions notwithstanding, it is important to underscore the fact that the scoring procedure used does render all findings in this study less definitive than would have been the case had an experimentally naive scorer been used.

[7] All P values reported in this paper are for the two-tailed test.

In summary, there was no evidence that *Fast* and *Fast-Distraction* conditions had differing effects on word association responses as scored in this experiment. None of the five indeterminate categories showed significant differences among the three groups. Of the eleven categories about which predictions were made, two showed significant *Slow-Fast* and *Slow-Fast-Distraction* differences.

EXPERIMENT II

In an attempt to clarify the rather ambiguous findings described above, a second experiment was conducted. Eighteen of the original 28 *Slow* group *S*s were retested individually on the same list and by the same *E*s 8–12 months after the original testing. In the hopes of effecting even briefer response times than were obtained under *Fast* and *Fast-Distraction* conditions, these *S*s were paid five cents for every response made within 1.20 seconds. So that the *S*s could pace themselves, the reaction time to each response was read aloud by *E* immediately after the response was given. Further, *S* also heard a click following each response made within the 1.20 second limit. As before, *S* was blindfolded and was given the usual 13 practice words prior to the test proper. Thus, this *Fast-Reward* group constituted 18 *Slow* *S*s retested under essentially *Fast* instructions with money serving as a reward for quick responding. As before, all responses were scored by the senior author.[8]

As was expected, *Fast-Reward* instructions led to significantly briefer latencies than did *Slow* instructions. In fact, every one of the 18 *S*s decreased in mean reaction time from *Slow* to *Fast-Reward* conditions, the group Mean and *SD* for the *Fast-Reward* latencies being 1.44 seconds and 0.11 seconds versus 2.31 seconds and 0.18 seconds for the same *S*s under *Slow* conditions. Contrary to expectation, however, response times under *Fast-Reward* conditions were not significantly briefer than the *Fast* and *Fast-Distraction* latencies (Table 1).

Word association responses under *Slow* and *Fast-Reward* conditions were compared, for each category, by means of the Wilcoxon Signed Ranks technique for paired replicates (5). In accord with prediction, there were significantly more *Perseveration*, *Perseveration-meaningful*, and *Clang* responses ($P < .01$) and significantly fewer *Subordinate* responses ($P < .01$) under *Fast-Reward* instructions than under *Slow* instructions. Of the indeterminate categories, *Attribute* showed a significant decrease in frequency from *Slow* to *Fast-Reward* conditions ($P < .01$). Table 3 shows Means and *SD*s for each category under test (*Slow*) and retest (*Fast-Reward*) conditions.

EXPERIMENT III

A third experiment was conducted in order to find out whether a group of schizophrenics would differ from a matched group of normals in the same way, with regard to word association responses, as normals under time pressure differ from normals without time pressure.

The *S*s for this study were schizophrenics and hospital aides from a Veterans Administration neuropsychiatric hospital. The patient group was a fairly typical sample of the testable members of this hospital's schizophrenic population. That is, most but not all of the sample were chronic patients of several years hospitalization; some were nearly free from overt symptoms and some were obviously psychotic. All patients were free from known organic impairment and none was undergoing electric shock treatment at the time of testing. The groups were matched, on a man-to-man basis, for age and verbal intelligence as estimated by the Vocabulary subtest of the Wechsler Adult Intelligence Scale. Group Means and *SD*s are reported in Table 4, none of the group differences approaching statistical reliability. The groups were not matched for edu-

---

[8] The scoring was identical to that of the previous experiment with one exception: because of certain scoring difficulties peculiar to *Multi-word-discrete*, this category was dropped from the classification system.

TABLE 3

CATEGORY SCORES FOR 18 *S*s TESTED UNDER *SLOW* INSTRUCTIONS AND RETESTED UNDER *FAST-REWARD* INSTRUCTIONS

Category	*Slow* Mean	*Slow* *SD*	*Fast-Reward* Mean	*Fast-Reward* *SD*
Immature[a]				
Completion	5.00	4.01	4.50	2.73
Distant	2.11	1.58	2.17	1.66
Perseveration	1.61	1.73	4.50	4.02
Perseveration-meaningful	6.17	3.55	8.83	4.24
Clang	0.50	0.79	1.44	1.51
Emotional	0.50	0.93	1.17	1.60
Repetition	0.78	1.23	0.67	0.85
Mature				
Synonym	5.28	2.58	4.78	2.42
Supraordinate	0.44	0.63	0.39	0.98
Subordinate	2.33	1.68	1.22	1.25
Indeterminate				
Attribute	5.94	3.47	3.44	1.92
Verb	3.89	2.62	3.22	1.57
Contrast-coordinate	13.67	5.82	15.39	4.96
Multi-word	1.44	1.10	1.22	1.34
Blocking	0.39	0.98	0.22	0.55

[a] The category *Multi-word-discrete* was not scored in this experiment. See Footnote 8.

TABLE 4

CHARACTERISTICS OF NORMAL AND SCHIZOPHRENIC GROUPS

Group	Age in Years		Verbal IQ		Grade Completed	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Normal ($N = 20$)	35.75	12.98	101.95	13.52	10.65	2.13
Schizophrenic ($N = 20$)	36.30	11.57	100.60	12.55	10.30	2.43

TABLE 5

CATEGORY SCORES FOR NORMAL AND SCHIZOPHRENIC GROUPS

Category	Normal Mean	Normal *SD*	Schizophrenic Mean	Schizophrenic *SD*
Immature				
Completion	2.00	1.49	1.50	1.91
Distant	1.45	1.19	6.00	9.10
Perseveration	0.10	0.32	1.70	2.55
Perseveration-meaningful	4.10	2.48	3.75	2.17
Clang	0.40	1.14	0.95	1.25
Emotional	0.30	0.73	0.30	0.48
Repetition	1.20	1.14	2.70	4.73
Mature				
Synonym	7.00	4.52	4.45	3.20
Supraordinate	2.95	2.79	2.90	2.87
Subordinate	1.80	1.46	1.15	1.25
Indeterminate				
Attribute	3.15	2.02	2.20	1.87
Verb	2.65	2.10	1.60	1.96
Contrast-coordinate	16.30	8.93	10.35	8.60
Multi-word	4.00	5.09	10.14	14.23
Blocking	3.95	5.53	7.50	8.69

cation, although group differences on this variable similarly turned out to be far from statistical significance. The word association test was administered individually to the *S*s of both groups by the fourth author. The instructions for the two groups were in all essentials identical to those previously given to the *Slow* group *S*s in Experiment I, i.e., the *S* was told to give his first associated word but not to feel rushed or hurried, etc. As was the case with the previous groups, the senior author scored all protocols.

The Wilcoxon Signed Ranks technique for paired replicates was again used to assess group differences for each category. Significant differences in the predicted direction were found for *Distant* ($P < .02$), *Perseveration* ($P < .01$), *Clang* ($P < .05$), *Synonym* ($P < .05$), and *Subordinate* ($P < .05$). No other group differences attained statistical reliability. Table 5 shows Means and *SD*s for each category by group.

In order to assess the possible effects of age, education, cultural background etc. on word association as scored in the present system, a second, incidental analysis was performed. The 12 *Slow* group males were compared with the 20 hospital aides on each category by means of the Mann-Whitney "U" Test for unmatched groups (5). The student group showed significantly more *Completion* ($P < .01$), *Perseveration* ($P < .01$), *Perseveration-meaningful* ($P < .05$), and significantly fewer *Blocking* ($P < .01$) and *Supraordinate* ($P < .01$). These findings, although possibly suggestive of further word association research within the framework of differential psychology, were in no sense predicted here and have no theoretical status in the present experiment.

DISCUSSION

The findings of these three experiments can be approached from two points of view: (*a*) as empirical information about the word association process in various groups and under different conditions of administration; (*b*) as findings relevant to the alleged microgenesis of word association. Considered from the first point of view, some of our data are relevant to previous association studies. Thus, Siipola, Walker, and Kolb (8) recently described an experiment comparing the word association responses of groups tested under free versus time pressure conditions. They found that noun responses to adjective stimulus words significantly decreased under time pressure conditions. This result is paralleled by the decrease in *Attribute* responses under *Fast-Reward* conditions in the present study (Experiment II). Siipola *et al.* also found, however, that "contrast" responses significantly increased under pressure instructions. The present study, on the other hand, failed to find any significant group differences for *Contrast*, although the category presumably had somewhat different content. Similarly, Rapaport, Gill, and Schafer (6) compared the word associations of normals and schizophrenics using a number of scoring categories, some of which are fairly comparable to those used in the present investigation. They too found a significantly higher frequency of categories roughly equivalent to *Distant*, *Perseveration*, and *Clang* in the schizophrenic group than in the normal group. Contrary to our own findings, however, Rapaport *et al.* also report significant group differences for the near-equivalents of categories *Repetition* and *Completion.* It is of course especially hazardous to try to interpret the *divergent* results of two such studies, in view of the lack of complete identity of the categories on which findings are compared.

More important within the present framework, however, is the possible significance of the results for a microgenetic view of word association. It will be recalled that the only categories for which predicted group differences were confirmed were the following: (*a*) Experiment I—*Perseveration* and *Distant*; (*b*) Experiment II—*Perseveration*, *Perseveration-meaningful*, *Clang*, and *Subordinate*; (*c*) Experiment III—*Perseveration*, *Distant*, *Clang*, and *Subordinate.* Perhaps a parsimonious interpretation would be as follows. Insofar as the findings are reliable (see again Footnote 6), a small group of categories—including *Perseveration* and perhaps *Distant*, *Clang*, and *Subordinate*—seems best to express word association response tendencies common to schizophrenics and to normals under time pressure. If the further interpretation is adopted that this common core of categories reflects a common attained level of microdevelopment in the associative process, then it seems very likely that the original hypothesis, involving a larger number of categories, was too broad and nonspecific. That is, it may well be that microdevelopmental immaturity, at least insofar as word association is concerned, is most accurately characterized by some but not all of the attributes that at one time or another have been assigned to it.

Such a view of the present results suggests several broad lines of future investigation. First of all, further attempts could be made to verify the present findings within the area of word association, with the added intent of specifying more closely the cognitive processes that lie behind the specific responses scored as *Perseveration*, *Clang* etc. Second, one could endeavor to develop methods of investigating microgenetically immature cognitive responses in different types of tasks, for example, the completing of incomplete sentences under time pressure. In such a study it would be of paramount interest to find out whether or not it would be precisely those categories analogous to *Perseveration*, *Clang* etc. which increase in frequency under pressure conditions. Further, it would be of great interest to determine whether this same response cluster emerges in normals under debilitating conditions other than time pressure, e.g., under the influence of various kinds of pharmaceutical agents.

Finally, it is important to add that, in our view, microgenetic concepts seem at present to constitute a somewhat specialized developmental approach to cognitive phenomena rather than a cognitive "theory" in any strict sense of the term. For example, the microdevelopmental formulation is hardly rigorous enough in its present form to rule out alternative explanations of the present findings cast in terms of the rigidifying effects of time pressure, the inability of schizophrenics to maintain a constant word association set, and the like. As an approach therefore, its chief value may ultimately lie in whatever fecundity it may possess in the way of generating fresh and novel experimentation on familiar, commonplace phenomena.

SUMMARY

The present study constitutes an attempt to test certain general hypotheses derived from a microgenetic approach to word association. In the first two experiments, using college students as subjects, association responses given under time pressure conditions were compared with those given without time pressure. In the third experiment, word associa-

tions of a group of schizophrenics and a matched group of hospital aides were similarly compared, both groups in this case responding without time pressure. It was predicted that the word associations of the college students performing under time pressure would differ from those of the college students responding under free conditions in exactly the same way as the responses of the schizophrenics would differ from those of the aides. Further, the precise character of the group differences in question was specified in the hypotheses. These predictions were in part supported by the results. That is, several of the predicted group differences were confirmed and a partial congruence between student-student differences and schizophrenic-aide differences was found. An interpretation of these findings was suggested and possible directions for further research within a microgenetic orientation were specified.

REFERENCES

1. ARIETI, S. *Interpretation of schizophrenia*. New York: Brunner, 1955.
2. FLAVELL, J. H., & DRAGUNS, J. A microgenetic approach to perception and thought. *Psychol. Bull.*, 1957, **54**, 197–217.
3. JUNG, C. G. *Studies in word association*. New York: Moffat, 1919.
4. KARWOSKI, T. F., & BERTHOLD, F. Psychological studies in semantics. II. Reliability of free association tests. *J. soc. Psychol.*, 1945, **22**, 87–102.
5. MOSES, L. E. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122–143.
6. RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing*. Vol. 2. Chicago: Year Book Publishers, 1946.
7. SCHILDER, P. On the development of thoughts. In D. Rapaport (Ed.), *Organization and pathology of thought*. New York: Columbia Univer. Press, 1951. Pp. 497–518.
8. SIIPOLA, ELSA, WALKER, W. NANNETTE, & KOLB, DOROTHY. Task attitudes in word association, projective and nonprojective. *J. Pers.*, 1955, **23**, 441–459.
9. WELLS, F. L. A preliminary note on the categories of association reactions. *Psychol. Rev.*, 1911, **18**, 229–233.
10. WILSON, K. V. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, **53**, 96–101.

Received June 18, 1957.

Bureau Ednl. Research
DAVID H[illegible] S COLLEGE
Dated ....
Accs. No ....

IMPROVEMENT IN THE PERFORMANCE OF SCHIZOPHRENICS ON CONCEPT FORMATION TASKS AS A FUNCTION OF MOTIVATIONAL CHANGE[1]

DAVID K. CAVANAUGH

University of Buffalo

IMPAIRMENT of performance on concept formation tasks has been demonstrated repeatedly as a characteristic of schizophrenia (1, 2, 12, 19, 20). A variety of interpretations have been offered as to the origin and nature of this impairment.

A group of theorists typified by Goldstein stress the similarity between the performance of schizophrenic and organic patients. Goldstein indicates that the similarities in the behavior of organics and schizophrenics points to an organic process (perhaps of secondary origin) as cause of schizophrenic impairment (9, 10, 11). Another group of theorists typified by Cameron (3, 4, 5, 6) suggest that the personality disorganization of the schizophrenic is basic to the latter's decrement in conceptualization. Thus, a high degree of abstracting ability may be present in the schizophrenic and be confounded by the severe personality disorientation. For Cameron this represents the product of the schizophrenic's "social disarticulation."

Whiteman (22) notes that from Cameron's theory one can predict a differential deficit with schizophrenics on socially toned as opposed to formally toned concepts. Goldstein, on the other hand, would seem to suggest a generalized deficit with no differential decrement as a function of social test content. Whiteman reports findings which he interprets as "lending presumptive support to a theoretical position which also stresses the importance of social withdrawal as a determinant of cognitive functioning in schizophrenia" (22, p. 271).

Another group of theorists emphasize the distinction between performance and potential ability with schizophrenics. Shakow (14, 18) distinguishes between schizophrenic learning ability and capacity. Hunt and Cofer (13) conceive of psychological deficit in schizophrenia as a concomitant of mental illness. They describe behavior as a product of the stresses faced by an individual and his abilities to handle these stresses. In schizophrenia, the patient's limited ability to handle stress leads to such abnormal behavior as withdrawal. The addition of a new stress, more potent than the stresses which have led to schizophrenia, may at least temporarily reverse the trend toward inappropriate behavior and result in actions more similar to those of normals. The psychological deficit of schizophrenics is thus perceived as a performance variable rather than an accurate indicant of the patient's present capacity. Pascal and Swenson (17) report results consistent with the position of Hunt and Cofer. Using an escape from aversive stimuli design, Pascal demonstrated schizophrenic performance at near the level of normals on a complex discrimination reaction time type of learning problem. Cohen (7) used an escape from aversive stimuli design on a task involving the learning of a series of motor responses to successively presented visual stimuli. His results indicated facilitation for the more motivated schizophrenics as compared to patients under mere rapport conditions.

The purpose of the present study was to investigate the role of motivation in the performance of schizophrenics on concept formation tasks. Two questions were asked: Does escape from aversive stimuli decrease impairment in concept formation with schizophrenics? And are there differences in the formation of social as opposed to formal concepts under this condition of increased motivation?

METHOD

Subjects

Hospitalized male veterans were used as *S*s in the study. The 36 normal *S*s were obtained from the medical and surgical wards of the VA General Medical and Surgical Hospital at Buffalo, N. Y. Patients were selected whose admission diagnosis and past medical history were reasonably devoid of psychosomatic etiology. The age range was from 20 to 40 years.

---

[1] This report is adapted from a thesis submitted in partial fulfillment of the requirements for the Ph.D. degree at The University of Buffalo. The author is indebted to Walter Cohen, under whose guidance the study was carried out, and to Ira Cohen for his advice and assistance. Grateful acknowledgment is also made of the criticisms and suggestions of others of the psychology faculty of The University of Buffalo. The writer is grateful to the professional staffs of the VA hospitals at Buffalo and Canandaigua, N. Y., for their generosity and interest in supplying the time and facilities necessary for the completion of the study.

The 108 schizophrenic *S*s were selected from the VA Neuropsychiatric Hospital at Canandaigua, N. Y. These individuals had been diagnosed and were being treated as schizophrenics. Only those patients were used whose diagnoses were uncomplicated by neurological involvement. As with the normals, the age range was from 20 to 40 years. For this sample, the additional criteria of contact and cooperativeness were necessary. No systematic differences were noted between subgroups with regard to such variables as age, type of schizophrenia, educational level, or vocational level. For both normals and schizophrenics, the following proportions of the groups obtained the Wechsler-Bellevue vocabulary weighted score units as specified: 13—6%; 12—12%; 11—28%; 10—33%; 9—18%; 8—6%. The average weighted score for both populations was 10.39.

Tests and Apparatus

In addition to the vocabulary subtest from the Wechsler-Bellevue, two concept formation tests were administered to each *S* in the study. These tests were essentially the same as those which Whiteman (22) found to be most effective in differentiating between schizophrenic and nonpsychotic groups. The formal concepts are defined by Whiteman as reflecting noninterpersonal stimuli which can be described in physical, psychophysical, or quantitative terms such as volume or number, or in logical-relational terms as with concepts based on principles of hierarchical classification (e.g., species, genus) or relational concepts such as "middleness." This test is composed of 11 items with each item consisting of five 3″ by 3″ cards. The *S*'s task is to place each set of five cards in a proper sequence according to a logical-relational concept such as arrangement according to size of the box shown in each picture (Item 1) or according to the period of architecture noted in each picture (Item 7). The items are scored for time and accuracy, with a maximum time of 90 seconds per item.

The social concepts are described by Whiteman (21) as representing abstractions common to a number of situations involving human interaction. This test is composed of 21 items, 18 of which contain four 3″ by 3″ stimulus cards, while three of the items contain six stimulus cards. In all items, three of the cards depicted instances of one social concept, whereas one (or three) of the cards did not depict this concept. In executing the social concept items, *S*'s task was to separate the three cards that had the same idea from those that did not have that particular idea. For example, in the "rescue" item (Item 1), the three concept cards included (*a*) a scene depicting a fireman rescuing a person from a burning building, (*b*) a picture of a man rescuing a girl from the path of an advancing truck, and (*c*) a picture of a man throwing a life preserver to a swimmer in distress. The nonconcept card was that of a scene showing two children on sleds racing toward a busy intersection. The items were scored for both time and accuracy, with a maximum time of 90 seconds for the four-card items and 120 seconds for the six-card items.

The apparatus which supplied the aversive stimuli described below consisted of a tape recorder and ear phones. The output of a white noise generator was recorded on a one-hour tape. The generator was set at an intensity level of 116 decibels above absolute zero (10^{-16} watts per square centimeter) with random frequencies from 60 to 10,000 cycles per second (limited by recording apparatus and earphones). This level of noise was that selected by Pascal and Swenson (17) in the previously described study and was reported by them as being disturbing but not injurious. When administered, the noise began at the time of presentation of a given test item and was terminated upon either the correct completion of the item by the *S* or the expiration of the maximum time limit for the item.

Procedure

On the basis of preliminary screening, interviews, and testing, a pool of available *S*s was established for each of the groups. From this aggregate, *S*s were assigned to experimental and control groups in a random manner, stratified according to intelligence level. The four groups may be defined as follows: 1. Normal—Control (NC) Group: 18 normal (nonpsychiatric) *S*s who received the concept formation tasks under the usual conditions for the administration of psychological tests. 2. Normal—Noise (NN) Group: 18 normals who received the tests in the presence of white noise which was terminated at the successful completion of each test item or the expiration of the maximum time limit for each item. 3. Schizophrenic—Control (SC) Group: 54 schizophrenics who received the tests under conditions similar to those of Group NC. 4. Schizophrenic—Noise (SN) Group: 54 schizophrenics who received the tests under conditions similar to those of Group NN.

The order of presentation of the tests was randomized throughout the testing of all samples. The instructions given to the *S* at the time of testing were those used by Whiteman (21), consisting of a brief description of the tests and the definition of *S*'s task as previously noted. For those *S*s who received the tests under the noise conditions, additional instructions were given. As in Pascal and Swenson's study (17), the *S* was told that the experimenter was interested in seeing how well *S* could perform under conditions of distraction. It was indicated that if a particular arrangement of items was the incorrect solution, the noise would continue until *S* formed the correct solution. Similarly, in the nonnoise groups, *S*s were informed of incorrect solutions and urged to rework the problems. On completion of a test item, *S* was asked to verbalize the idea present in his arrangement of the cards. Inability to verbalize a realistic interpretation of the concept resulted in the item being scored as though it had not been completed.

Two scoring systems were used in the analysis of the *S*s' performance on the tasks. The first scoring system (S) gave three credits for an appropriate response within the time limit, regardless of the number of corrections made. The second scoring system (B) added time bonus credits to the procedure of Scoring System S. System S was intended to reflect whether or not the concepts were available to the individuals, whereas System B was directed toward the problem of how easily the concepts were available.

RESULTS

An A × B × L analysis of variance design (15) was used in the present study. "A" represents the population variable (schizophrenic and normal), and "B" signifies the experimental conditions (noise and nonnoise). "L" represents the control variable of intelligence introduced to achieve higher precision in the evaluation of A and B. Six of these analyses are included in the present report, representing scores of the *S*s to each of the tests by itself as well as total scores (social plus formal) for each *S*. The scores were computed by means of Scoring Systems S and B described above. A relatively low correlation was found between Systems S and B ($r = .28$).

The results of the six A × B × L analyses are summarized in Table 1. In all cases, the SN group approximated the performance level of the normals, and the SC group performed at a level significantly lower than the other three. On the formal test (Scoring System B), the low performance of the SC group was offset by the very high performance of the SN group, so a significant difference between psychotic and normals was not obtained.

In all analyses, the variable of noise versus nonnoise was found to display significant effects, as was the case with the control variable of intelligence. The interaction of conditions and populations was significant in four of the six analyses. In the case of the social and total scores with Scoring System S, this interaction was not significant. In these cases, the effects of noise with the schizophrenics were not as pronounced as in the other analyses, although the schizophrenic groups performed at significantly different levels, and the schizophrenic noise group was not significantly differentiated from the normals.

TABLE 1

SUMMARY OF ANALYSES OF VARIANCE FOR TOTALS, SUBTESTS, AND SCORING SYSTEMS

Source		Scoring System B (with time bonus)			Scoring System S (without time bonus)		
		Total	Social	Formal	Total	Social	Formal
A	Schizophrenic vs. Normal	.05	.05	—	.001	.001	.05
B	Noise vs. Control	.001	.001	.001	.001	.005	.005
AB	Interaction	.005	.01	.05	—	—	.05
L	Intelligence	.001	.001	.001	.001	.001	.001
AL	Interaction	—	—	—	—	—	—
BL	Interaction	.05	—	.005	.05	—	.01
ABL	Interaction	—	—	—	—	—	—

Note.—Numbers refer to the respective significance levels.

The populations × intelligence interaction was not significant in any of the analyses, as was the case for the populations × conditions × intelligence interaction. In two cases (the social test in both scoring systems), the conditions × intelligence interaction was not significant. Inspection of the data in the other four analyses revealed that this interaction was significant with the formal test, which resulted in significance also being found for the total scores with this interaction. Evaluation of the formal test data suggested that although the effects of noise were not consistent at all intelligence levels, no systematic trends could be found to account for this lack of consistency.

DISCUSSION

The schizophrenics performed at a level significantly inferior to normals under the usual testing conditions. These results indicate that the concept formation tests developed by Whiteman are of value in differentiating between normals and schizophrenics. The schizophrenics who took the tests in the presence of the white noise which was terminated upon the formation of correct responses consistently performed at a level superior to that of the schizophrenics under usual conditions. In fact, they approximated the functioning of the normal groups. This finding serves to re-emphasize the importance of the distinction between samples of performance and inferences about capacity levels with schizophrenics. This result is consistent with that reported by Cohen (7) and by Pascal and Swenson (17) and the theoretical position of Hunt and Cofer noted previously. The finding that the normal groups were not differentiated by the motivation conditions is also reported by Cohen. His interpretation is that normals work near their limit of proficiency under ordinary testing conditions.

Under conditions of increased motivation, schizophrenics attained the performance level of normals on both social and formal test material. The similarities between the analyses based on the two scoring systems suggest that the concepts were not only available to the schizophrenics but seemingly were as easily available to them as to the normals. Increased decrements with schizophrenics on social as opposed to formal concepts, as reported by Whiteman (22), were not apparent for the SN group of the present study. Within the present design, it was not feasible to compare directly the performance of the SC group on the two types of task content.

The results of the present study do not support the relative permanence and irreversibility of cognitive decrement indicated by Goldstein and other organically oriented theorists. The results also fail to support the differential deficit and progressive decrement in capacity for appropriate thinking implied by Cameron (3, 4, 5, 6). It would appear that normal, socially acceptable thinking may be relatively easily available to the schizophrenic, at least for schizophrenics who meet the criteria of contact and cooperativeness necessary for the present study.

The improvement in schizophrenic performance under the conditions of the present study refers only to a relatively brief and temporary improvement in a limited area of cognitive functioning. If subsequent research bears out the present findings in other areas of functioning and with more disoriented patients, a variety of implications are evident. In the treatment of schizophrenia, for example, it may be that efforts to modify the motivation of schizophrenics will prove more fruitful than the more usual attempt to retrain the patient's disorganized thinking.

The suggestion that schizophrenics suffer from decreased motivation, evident in the present study, is present in several theories of abnormal behavior. Theorists such as Dollard and Miller (8) and Mowrer (16) are representative of writers who describe personality maladjustment in terms of the patient's unsuccessful attempts to cope with anxiety. Within this framework, aspects of schizophrenic behavior are said to connote defensive efforts oriented around the attempt to reduce anxiety. Withdrawal from anxiety-arousing situations, avoidance of the stresses of adult environmental demands and a retreat from reality, become useful media for reduction of anxiety in the schizophrenic. In the absence of more appropriate methods of coping with anxiety, the schizophrenic's adjustment appears relatively more satisfactory to him than does his former adjustment. Changes in his present motivation toward relinquishing schizophrenic defenses would thus appear as a prerequisite for a return to a more adequate adjustment.

In addition to research which might help to define the limits of hypotheses noted in the present report and to explore the generality of the present observations, investigations of motivation as related to avoidance learning or reward learning paradigms, for example, might lead to even more emphatic effects on schizophrenic behavior.

Summary

This study tested the effects of escape from aversive stimuli conditions upon the performance of schizophrenics and normals. Concept formation tasks involving social and formal concepts were used as measures of performance. The schizophrenics and normals were each divided into two groups, Noise and Control, and matched on an intelligence measure. The concept formation tests were then administered. The control groups received the tests under the usual conditions of psychological testing. White noise was introduced through earphones to the *Ss* of the Noise groups at the presentation of each test item. The noise continued until the item was completed or the maximum time limit of the item was exceeded.

Under the conditions of escape from aversive stimuli, schizophrenics approximated the performance level of the normals with both types of test content. Schizophrenics who took the concept formation tests under the usual conditions of psychological testing performed at a level both inferior to and significantly differentiated from the normals and the more motivated schizophrenics.

The findings were interpreted in terms of an experimentally induced increase in schizophrenic motivation, accompanied by a temporary relinquishing of schizophrenic defenses and consequent performance at a level more representative of potential ability than that elicited in the usual testing situations. Suggestions for further evaluation of the role of motivation in schizophrenia were made.

REFERENCES

1. Benjamin, J. D. A method for distinguishing and evaluating formal thinking disorders in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. of California Press, 1946. Pp. 65–90.

2. Bolles, M. M., & Goldstein, K. A study of the impairment of "abstract behavior" in schizophrenic patients. *Psychiat. Quart.*, 1938, **12**, 42–65.
3. Cameron, N. Reasoning, regression, and communication in schizophrenics. *Psychol. Monogr.*, 1938, **50**, No. 1 (Whole No. 221).
4. Cameron, N. Deterioration and regression in schizophrenic thinking. *J. abnorm. soc. Psychol.*, 1939, **34**, 265–270.
5. Cameron, N. Schizophrenic thinking in a problem-solving situation. *J. ment. Sci.*, 1939, **85**, 1012–1035.
6. Cameron, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. II. New York: Ronald, 1944. Pp. 861–921.
7. Cohen, B. D. Motivation and performance in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **52**, 186–190.
8. Dollard, J., & Miller, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
9. Goldstein, K. The significance of special mental tests for diagnosis and prognosis in schizophrenia. *Amer. J. Psychiat.*, 1939–1940, **96**, 575–578.
10. Goldstein, K. Methodological approach to the study of schizophrenic thought disorder. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. of California Press, 1946. Pp. 17–40.
11. Goldstein, K., & Scheerer, N. Abstract and concrete behavior. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
12. Hanfmann, Eugenia, & Kasanin, J. Conceptual thinking in schizophrenia. *Nerv. ment. Dis. Monogr.* 1942, No. 67.
13. Hunt, J. McV., & Cofer, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. II. New York: Ronald, 1944. Pp. 971–1032.
14. Huston, P. E., & Shakow, D. Learning in schizophrenia. *J. Pers.*, 1948, **17**, 52–74.
15. Lindquist, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
16. Mowrer, O. H. *Learning theory and personality dynamics*. New York: Ronald Press, 1950.
17. Pascal, G. R., & Swenson, C. Learning in mentally ill patients under conditions of unusual motivation. *J. Pers.*, 1952, **21**, 240–249.
18. Shakow, D. The nature of deterioration in schizophrenic conditions. *Nerv. ment. Dis. Monogr. Serv.*, 1946, **70**, 1–88.
19. Vigotsky, L. Thought in schizophrenia. *Arch. Neurol. Psychiat.*, 1934, **31**, 1063–1077.
20. Wegrocki, H. Generalizing ability in schizophrenia. *Arch. Psychol.*, 1940, No. 254.
21. Whiteman, M. The performance of schizophrenics on social concepts. Unpublished doctoral dissertation, New York Univer., 1952.
22. Whiteman, M. The performance of schizophrenics on social concepts. *J. abnorm. soc. Psychol.*, 1954, **49**, 266–271.

Received July 3, 1957.

SOME CLINICAL CORRELATES OF OPERANT BEHAVIOR[1]

MARTHA T. MEDNICK AND OGDEN R. LINDSLEY

Harvard Medical School

When techniques of operant conditioning are adapted to investigate the behavior of chronic psychotic patients (5, 6), the subject (*S*) is placed in a small room equipped with a lever which can be manipulated in order to obtain various types of reinforcements. Appropriate brief verbal instructions are given. The intensive study of approximately 50 *Ss* recently revealed that there are great individual differences in the rate at which a patient pulls a lever in order to obtain a particular reinforcement. In searching for an explanation of these differences, it has been found that the rate of response is not related to admission diagnosis, intelligence quotient, age, or total time of hospitalization (10). On the basis of an inspection of the data and an extremely gross rating of patients' clinical behavior, however, the rate of response appeared to be directly related to "depth of psychosis or severity of illness" (10). An attempt to verify this relationship by King *et al.* (2) with *acutely* ill psychotic patients was in the predicted direction but not statistically significant. They observed a concave downward curvilinear relationship between severity of illness and rate of operant responding. The group evidencing medium severity of neuropsychiatric illness responded at the highest operant rate.

The present study was primarily designed to examine systematically the relationship of severity of illness to experimental performance in a chronic patient population.

METHOD

Twenty two male chronic psychotic patients, hospitalized from 3 to 47 years with a median of 16 years, and six male hospital attendants were the *Ss* used. These *Ss* were all taking part in an ongoing study in operant conditioning. The patient group had had from 10 to 448 hour long sessions of experience in the experimental room, with a median of 18 sessions. Each normal was run for 50 hours and was then placed on extinction. Briefly, the operant conditioning session consisted of placing each *S* alone for one hour in a small room equipped with a vending machine device. Figure 1 depicts this experimental situation. The *S* was given verbal instructions to the effect that if he operated the lever he would get candy (or some other reinforcement). A detailed statement of method is given by Lindsley (5). The *Ss* in this study were reinforced with a mixture of penny candy and cigarettes. The reinforcements were delivered on a one-minute variable interval schedule (1′VI), i.e., an average of one reinforcement per minute.

The Lucero-Meyer Fergus Falls Behavior Sheet (LMBS) (7) was chosen to rate the ward behavior of the patients. This scale is brief and simple, dealing with observable hospital behavior. It has been used with a chronic population and seems sensitive to differences within such a group while being independent of psychiatric diagnosis. In addition, changes in a patient's behavior can be picked up over a short length of time. The authors of the scale report reliability coefficients of .92 and .94. In the current study, no further check of reliability was made.

A brief diagnostic test battery was also given. A short form of the Wechsler-Bellevue, Form I, consisting of Vocabulary, Comprehension, Block Design, and Picture Completion (C-BD-V-PC) was administered. The C-BD-V-PC was found by Patterson to correlate .96 with the full scale WB-I on a heterogeneous psychiatric population (8). If *S* was untestable by this form of the Wechsler-Bellevue, the Ammons Full Range Picture Vocabulary Test (APV), Form A, was attempted. In addition to the intellectual measure, the Rorschach was administered. The final clinical index obtained was the Tulane Psychological Test Behavior Rating Scale (TBS) developed by King (3). This scale represents an attempt to quantify the quality of behavior observed during psychological testing. The TBS was filled out by the clinical psychologist following each testing session. The operant behavior data were not known to the clinical psychologist until the patients had all been tested.

RESULTS

Operant Data

Three response measures were used in the analysis of the operant conditioning data: the rate of response (R Hr.), the total number of inter-response times greater than 10 sec. (IRT $> 10''$), and the sum of the inter-response times greater than 10 sec. ($\sum$ IRT $> 10''$), i.e., the total amount of time during which *S* did not respond. These indices are described in detail by Lindsley (4). The latter

---

[1] The work reported in this paper was accomplished under Contracts N5 ori 07662 and Nonr 1866(18), sponsored by the Group Psychology Branch, Office of Naval Research and under research grant MH 977 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service. This paper was written while the senior author was a USPHS Post-doctoral Research Fellow at the Harvard Medical School.

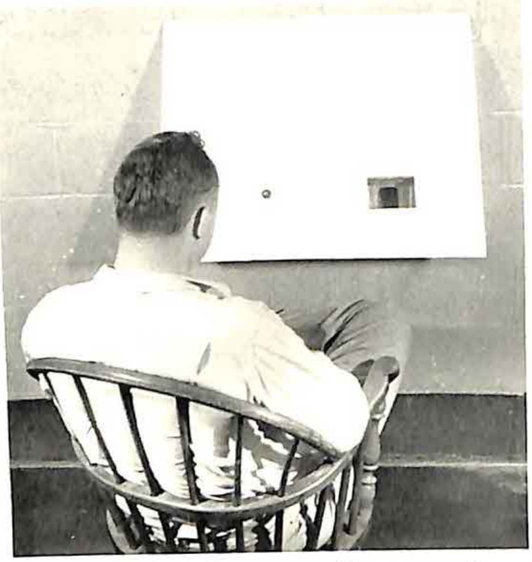

FIG. 1. STANDARD APPARATUS USED FOR OPERANT CONDITIONING OF PSYCHOTIC PATIENTS; *S* PULLS THE KNOB ON THE LEFT OF THE PANEL AND IS REINFORCED BY THE INTERMITTENT PRESENTATION OF CANDY AND CIGARETTES IN THE MAGAZINE CHUTE ON THE RIGHT (DESCRIBED IN DETAIL IN [6] AND [4])

two measure intra-hour variability. The present paper will be concerned in the main with R/Hr. data.

Table 1 summarizes the operant conditioning data for testable and untestable patients and the normal *Ss*. The data presented represent the median of the first ten hours in the study for each *S* in each group. It should be noted that the # IRT $>$ 10 sec. and the $\sum$ IRT $>$ 10 sec. must be compared in relation to each other. Thus, the untestable patients have few pauses but do not respond for almost the entire hour, while the testable patients take numerous short breaks. The normals only occasionally take a short break. It is clear that the two patient groups differ from each other as well as from the normals with respect to responsivity and intra-hour variability.

TABLE 1

OPERANT CONDITIONING RESPONSE MEASURES FOR TESTABLE, UNTESTABLE, AND NORMAL *Ss* (median values)

	R/Hr.	# IRT 10″	IRT 10″ (in min.)
Untestable	19	9	58
Testable	1,421	44	41
Normals	9,566	2	0

Note. Statistical evaluation of the differences between the three groups were made by the Finney 2 × 2 contingency tables. All groups differed significantly ($p < .01$) on median $\sum$ IRT $>$ 10″. The difference between testable and untestable *Ss* on median R/Hr. was significant ($p < .04$).

TABLE 2

TEST AND RATING DATA DESCRIBING UNTESTABLE, TESTABLE, AND NORMAL GROUPS (median values)

	N	Yrs Hosp.	IQ[a]	Ratings		Rorschach			
				TBS	LMBS	*R*	*M*	W	*F*+ %
Untestable	12	14	—	21	1.9	—	—	—	—
Testable	10	18	84	306	3.1	9	—	9	86
Normal	6	—	107	420	—	13	2	8	95

[a] IQ for testable patient *Ss* was on basis of abbreviated WB I or APV, Form A.

Behavior Ratings

The LMBS consists of 11 short scales, each pertaining to a particular kind of hospital behavior. Within each scale are five descriptions of behavior, ranging from most to least disturbed. The rater is instructed to place a check mark next to the description which best describes the patient. Each description receives a numerical score from *one* for the most disturbed type of behavior to *five* for the least disturbed. The scores for all scales are averaged to obtain the total LMBS rating. There is thus a possible range of 1.0 to 5.0. The LMBS ratings of the patient group in this study ranged from 1.3 to 3.8 with a median of 2.3. These values are comparable to the mean score of 2.5 and a range of 1.3 to 4.7 reported by Lucero and Meyer in their account of construction of the scale (7).

With respect to the relation of LMBS ratings to operant behavior, the median rate of response for the first ten hours was found to have a rank difference correlation of .23 (not significant) with the median LMBS ratings. The rank difference correlation between the LMBS rating and the median rate of response for the ten hours nearest to the date of the rating, however, was .81, significant at the .001 level.

Individual Testing

Individual testing was attempted with all the patients in the group. Twelve of these patients were completely untestable. Six C-BD-V-PCs were administered and four APVs. All of the ten testable patients were able to perform at least the free association to the Rorschach. The test data, presented in Table 2, reflect the extreme chronicity of the population. The test data for the normal *S*s are also presented.

The Test Behavior Rating Scale (TBS) was scored by the tester. It consists of 12 subscales, each having a possible range of scores of 0 to 40. A final score is obtained by summing all subscale scores so that the total score can range from 0 to 480. A score of 480 would imply a very good adjustment to the testing situation. The TBS ratings for the patients ranged from 0 to 336, with a median of 95. The original samples reported by King (3) in his development of the scale received mean scores from 299 to 399. It would seem, then, that the present population as a whole was more disturbed and less accessible than the Tulane patients.[2]

The rank difference correlation of the TBS and R/Hr. of the first 10 hours was .13, while R/Hr. for the 10 hours nearest to the date of testing correlated .39 with TBS. Neither of these coefficients is significant, although it should be noted that the correlational magnitude increased in the same general manner as in the case of the LMBS ratings. A significant direct relationship ($p < .04$) was found between sheer testability and rate of response, using the Finney 2 × 2 contingency tables (1). Similarly, high and low scores on the TBS were significantly related to R/Hr. (Finney contingency tables, $p > .019$). This is essentially a repetition of the testability-untestability findings, since high TBS scores were the testable patients and low scorers the untestable ones. A chi-square test was also performed on each subscale of the TBS. Two of the twelve subscales showed a significant relationship with the rate of response near testing date. These two were "ability to follow instructions" and "willingness to accept the task as an activity." The former verifies the impression, stated in an earlier project report, that those patients who are able to understand simple directions are also high operant responders (10). The subscales of the LMBS were similarly analyzed, but none of these proved significant.

DISCUSSION

This study strongly suggests that the operant conditioning performance of chronic psychotic patients and certain clinical variables are related. The earlier impression of a positive relationship between rate of response and severity of illness was supported. It is meaningful to look at the results in terms of a notion of adaptability to the limited social and physical demands of the hospital environment. All of the patients in the present sample can be described as severely ill; the length of hospitalization bears this out. While they all require institutionalization, they nevertheless differ widely in their ability to adjust within such a setting. The LMBS may be considered an index of this ability, and its significant relationship to operant conditioning may be reflective of a general adjustive ability. Individuals (in a chronic group) who are high operant responders may be those who remain sensitive to rewards in the social environment and can learn to manipulate this environment so as to obtain some of these rewards. The relationship of operant performance to clinical testability further supports this notion. Again, the high operant responders are those who can adjust to a type of environment demand.

The fact that the correlations of clinical ratings and operant responsivity are raised by utilizing R/Hr. nearest to the date of ratings is of interest. This would seem to reflect sensitivity of operant performance to clinically observed changes in the patient.

SUMMARY

Twenty-two chronic psychotic patients, *S*s in an operant conditioning study, were given psychological tests and rated as to their ward behavior. Ratings of ward behavior and rates of operant response were directly related. Those patients who were testable by at least one clinical test were those who were high operant responders. These findings were discussed in terms of a notion of adaptability to the demands of the hospital environment.

Clinical and operant data on six normal *S*s were also presented.

REFERENCES

1. FINNEY, D. J. The Fisher-Yates tests of significance in 2 × 2 contingency tables. *New statistical table No. VII*. Cambridge, England: Cambridge Univer. Press, 1948.
2. KING, G. F., MERRELL, D. W., LOVINGER, E., & DENNY, M. R. Operant motor behavior in acute schizophrenia. *J. Pers.*, 1957, **25**, 317–326.

---

[2] A communication with King has indicated that his ratings of detailed descriptions of our patients coincide with those assigned by the present investigators.

3. KING, H. E. *Psychomotor aspects of mental diseases.* Cambridge, Mass.: Harvard Univer. Press, 1954.
4. LINDSLEY, O. R. Annual Technical Report 3 to the Group Psychology Branch, Office of Naval Research. New techniques of analysis of psychotic behavior. Sept. 1955–Nov. 1956.
5. LINDSLEY, O. R. Operant conditioning methods applied to research in chronic schizophrenia. *Psychiat. Res. Repts.* 1956, **5**, 118–139; 147–153.
6. LINDSLEY, O. R., & SKINNER, B. F. A method for the experimental analysis of the behavior of psychotic patients. *Amer. Psychologist,* 1954, **9**, 419–420. (Abstract)
7. LUCERO, R. J., & MEYER, B. F. A behavior rating scale suitable for use in mental hospitals. *J. Clin. Psychol.*, 1951, **7**, 250–254.
8. RABIN, A. I., & GUERTIN, W. H. Research with the Wechsler-Bellevue test: 1945–1950. *Psychol. Bull.*, 1951, **48**, 211–241.
9. SKINNER, B. F. What is psychotic behavior? In *Theory and Treatment of the Psychoses: Some Newer Aspects.* St. Louis: Washington Univer. Studies, 1956.
10. SKINNER, B. F., SOLOMON, H. C., & LINDSLEY, O. R. Status Report II to Group Psychology Branch, Office of Naval Research. Final report for the first year on the study of psychotic behavior. June 1953–May 1954.

Received February 25, 1957.

THE EFFECT OF PERSON-GROUP RELATIONSHIPS ON CONFORMITY PROCESSES[1]

JAY M. JACKSON AND HERBERT D. SALTZSTEIN
University of Michigan

MANY previous studies of conformity behavior—among them the well-known experiments of Sherif and Asch (1, 2, 11)—have employed situations where the forces to conform were essentially cognitive in origin. A person in a situation of judgment or choice is confronted with contradictory information from two different sources. He has to weigh the evidence provided by his own perception of the stimulus against his knowledge of the actions or judgments of other persons. Festinger (5) has pointed out that a person will have a *need for social reality*, i.e., a need to depend upon information provided directly or indirectly by others, to the degree that his information from so-called physical sources is inadequate.

Forces to conform which are created by a person's need for social reality have their source in his desire to make an appropriate rather than an inappropriate response, or to perceive the world accurately rather than inaccurately. This process of coming to cognitive terms with his environment can be distinguished from other processes of conformity that are influenced by the person's membership or nonmembership in a group, the strength of his attraction to membership, or the relevance of the situation to the goals of the group.

In experiments by Festinger, Thibaut, Schachter, and others (5, 6, 10), forces to conform were created which derived from pressures towards uniformity in a problem-solving group. These forces, generated by a process called *group locomotion*, are induced upon all persons who belong to the group and, especially, upon any deviant who may be blocking the group's progress toward its goal. The more attractive a group is for a member, the stronger are the forces from this source acting upon him to conform (3, 5).

Forces to conform that arise out of the necessity for group locomotion clearly depend upon the existence of a group of interdependent persons who reciprocally expect each member to help rather than hinder the group. These forces should exist only in areas that are relevant to the group's goal-achievement. They will be induced upon all those accepted as members of the group, but not upon nonmembers. They will be strongest for those members who are most identified with the group and its objectives.

This paper reports an experiment designed to test hypotheses derived from these theoretical assumptions about *social reality* forces and *group locomotion* forces to conform.[2]

Normative and Modal Situations

Implicit in the previous discussion is a distinction between two types of situation in which conformity behavior can occur. A *normative* situation is one where group members are interdependently working towards a common goal, and consensus therefore takes on a prescriptive or normative value for members. A *modal* situation is one where either there is no common goal, or the task is not relevant to group achievement. Thus any agreement of opinion or judgment is not the result of reciprocal patterns of influence, but a mere coincidence of events, a piling up of responses at the mode.

Group Membership and Person-Group Relationships

Let us consider four types of relationship that occur between a person and a group:[3]

1. *Psychological Membership* (high *acceptance*, high *attraction*): The person is highly accepted as a member of the group by other members. They expect him to adhere to a member's role, conform to the norms of the

---

[1] This study was conducted under Contract Nonr-1224(11) with the Office of Naval Research. A mimeographed report containing the detailed instruments and procedures is available elsewhere (9). This article is based in part on a report to the annual meetings of the American Psychological Association, September 1956.

[2] This distinction is similar to that made by Deutsch and Gerard, who speak of *informational* and *normative* social influences upon an individual's judgment (4).

[3] It is pointed out elsewhere (8) that these person-group relationships are not discrete classes, but segments of a space.

group, and contribute to its goals. The person is motivated to take the role of member, thus highly attracted to membership in the group. He is both treated by others like a member of the group, and desirous of being a member.

2. *Psychological Nonmembership* (low *acceptance*, low *attraction*): The person has little or no acceptance as a member by other members, for whatever reason. He also has little or no attraction to membership in the group. Under these conditions he does not "belong" psychologically to the group.

3. *Preference Group Relationship* (low *acceptance*, high *attraction*): In this relationship, the person has little or no acceptance as a member by the other members. But his attraction to the group is highly positive, implying that he perceives the group as capable of fulfilling, directly or indirectly, some of his needs. He would prefer to belong to the group, but is not a member.

4. *Marginal Group Relationship* (high *acceptance*, low *attraction*): The other members of the group accept the person as a member. But he has little or no attraction to membership. He is considered to be a member by the others, but he is not motivated to take the role of member.

Summary Statement of Theory

The discussion has dealt with forces to conform which have their source in the need for *social reality* and in the necessity for *group locomotion* in a goal-oriented group. It is possible to summarize the reasoning in the form of two assumptions that specify the relative magnitude of these forces to conform that will act upon a person in each of the four person-group relationships, in both normative and normal situations:

I. *Forces to conform arising out of a need for social reality will have equal magnitude in both normative and modal situations, and will not vary with the person-group relationship.*

II. *Forces to conform arising out of the necessity for group locomotion will exist only for positively accepted persons in a normative situation, and will be positively related in magnitude to such persons' attraction to the group.*

In deriving hypotheses from these statements, we further assumed that forces to conform which act upon a given person in a particular situation are additive, and that his conformity behavior is determined by the sum of all forces to conform acting upon him in that situation.

TABLE 1

SOURCES AND MAGNITUDES OF FORCES TO CONFORM FOR FOUR PERSON-GROUP RELATIONSHIPS UNDER TWO TASK CONDITIONS

Person-group Relationship	Experimental Conditions	Normative Situation	Modal Situation
Psychological Membership	(high attraction, high acceptance)	social reality + high group locomotion (a)	social reality (e)
Marginal Group Relationship	(low attraction, high acceptance)	social reality + low group locomotion (b)	social reality (f)
Preference Group Relationship	(high attraction, nonacceptance)	social reality (c)	social reality (g)
Psychological Nonmembership	(low attraction, nonacceptance)	social reality (d)	social reality (h)

Hypothesis 1. *Conformity behavior of Members*[4] *will be greater in magnitude in a normative than in a modal situation.* The derivation of this prediction is extremely simple: social reality forces to conform acting upon Members have equal magnitude in both normative and modal situations (from I), but group locomotion forces act upon Members only in a normative situation (from II). The total force to conform acting upon Members will be greater, therefore, in the normative than in the modal situation.

Hypothesis 2. *Conformity behavior of Marginals will be greater in magnitude in a normative than in a modal situation.* The derivation of this and other hypotheses will be apparent in Table 1, where are summarized the relative magnitudes of forces to conform which act upon persons in the four person-group relationships and two situations, as postulated in our two basic assumptions.

Using the letters in this table for convenient designation, the first hypothesis can also be stated: conformity behavior in (*a*) > (*e*), or

$$[1] \qquad CB_a > CB_e$$

---

[4] Persons in the four relationships will be referred to as: *Members*, *Marginals*, *Preferences*, and *Nonmembers*.

The second hypothesis, similarly, can be stated:

$$[2] \quad CB_b > CB_f$$

By making comparisons between pairs of cells in Table 1, eleven other hypotheses,[5] summarized below, were derived from our two basic assumptions and tested in this experiment.

$$[3] \quad CB_a > CB_b$$
$$[4] \quad CB_a > CB_f$$
$$[5] \quad CB_b > CB_e$$
$$[6] \quad CB_a > CB_c$$
$$[7] \quad CB_a > CB_d$$
$$[8] \quad CB_a > CB_g$$
$$[9] \quad CB_a > CB_h$$
$$[10] \quad CB_b > CB_c$$
$$[11] \quad CB_b > CB_d$$
$$[12] \quad CB_b > CB_g$$
$$[13] \quad CB_b > CB_h$$

METHOD

Subjects

The *S*s were 100 male undergraduate volunteers, randomly assigned to 21 groups of four or five persons. During or immediately after the experiment, 14 of the *S*s indicated they had seen through some of the deceptions. The analysis is based upon data from the remaining 86 *S*s.

Experimental Design

In one half of the groups, *S*s were in a *high attraction* condition and in the other half a *low attraction* condition. In each group there was one *S* who was not accepted as a member by the others, each of whom was highly accepted by all. Thus, *high acceptance* and *nonacceptance* conditions were created. Each experimental session consisted of two periods, one under *normative* conditions and one under *modal*, with the normative always coming first. It would have been preferable to control on this sequence, but it proved impracticable to begin with the modal condition: this involved having a person unacceptable as a member of a "group" in which everyone works on his own. In each period, an experimental task was performed by the *S*s and data obtained about their conformity behavior. Questionnaires to validate the experimental conditions and to obtain additional data were administered immediately before and after each experimental period.

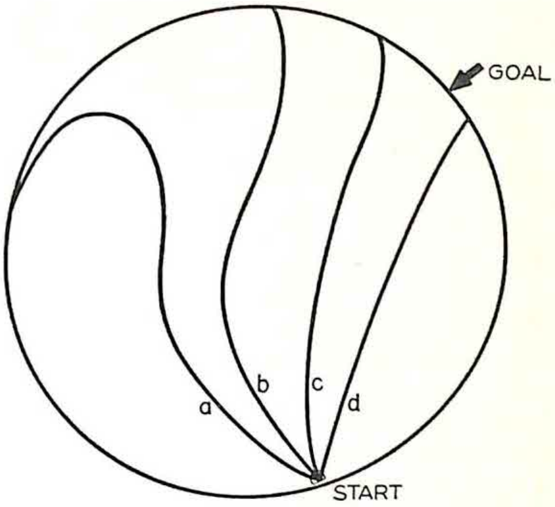

FIG. 1. A "SOCIAL VISION" PROBLEM; SUBJECT MUST JUDGE WHICH IS THE SHORTEST PATH FROM "START" TO THE "GOAL"

Experimental Task

The task consisted of a set of 10 "social vision" problems, each a variation of the problem seen in Fig. 1. This maze-like figure is a somewhat more complicated version of Asch's line problem (1, 2). *S*s were instructed to find the shortest path from the starting point to the black arrow. The experimenter pointed out that in each problem it became a choice between two of the four possible paths. By varying the position of the arrow around the circumference of the circle, 26 problems were constructed which varied in their degree of difficulty or ambiguity. Each set of 10 problems included a range from one extreme to the other, and the sets were matched for difficulty.

Each problem was projected onto a screen about 20 feet distant from the *S*s, who were seated in a semicircle. After a problem had been exposed for three seconds, *S*s were asked to write unsigned notes to other *S*s. Each note had to include the sender's opinion as to the correct answer and how confident he was of his response. He could also send any remarks he felt like writing. The notes were collected for "sorting" and after a suitable time interval, substitute notes were delivered, with each *S* receiving identical notes. On eight of the ten trials, the notes reported back the incorrect answer, held with a high degree of confidence by each other member of the group. After *S*s had read their notes, the same problem was then exposed for a second time. This time *S*s were asked to write their answers upon a special form provided for the purpose.

Inducing the Experimental Conditions

In the *high attraction* condition, *S*s were told that they had been matched for congeniality and ability to work well together, using information from their volunteer forms. The importance of the study was stressed to them. They were told about cash prizes that could be won. They were also told that each would have an opportunity later to learn more about the study and its results. In the *low attraction* condition, *S*s were told that we had been unable to match them, that the study was relatively unimportant, that they would not have an opportunity to learn about the research, and that prizes were no longer being awarded.

---

[5] Since all thirteen hypotheses are derived from two theoretical assumptions, they are clearly not independent. No attempt will be made, however, to aver the truthfulness of our theory by assessing the probability of achieving a given number of significant confirmations.

The acceptance conditions were induced by the following procedure. The experimenter announced that he had one more person than was needed, since all groups had to be a standard size. This had happened several times previously, so a procedure had been developed to eliminate one person "in a way that will be fair to all." *S*s were given a preliminary set of "social vision" problems, and spurious scores were announced, with all except a randomly selected person obtaining uniformly high scores, and the latter receiving a much lower score. *S*s were then asked to indicate on a ballot which person they wished to exclude. The experimenter, after scanning the ballots, announced that the randomly selected person was the person to be excluded. The nonaccepted person, who usually rose to leave at this point, was then asked by the experimenter to remain and participate in the task, because "later on everyone will be working on individual tasks, and we want you to have the same experience and practice as the others." The effect of this deception was to have a number of interdependent *S*s, who thought of themselves as a group about to work for a common goal (in the normative condition), and one other *S*, physically located in the group, engaging in the same activities but psychologically not accepted by the others as a member. Thus, in each group there were a number of *S*s in the *high acceptance* condition, and one *S* in the *nonacceptance* condition.

The *normative* condition was created by telling the "group members," i.e., the accepted persons only, that they would be given a group score on the task to follow, and would be compared as a group with other groups. The *modal* condition was created by telling all *S*s that each would be given an individual score on the task to follow and would be compared with all the other persons in his "position." (Each *S* sat before a small table. These "positions" were lettered consecutively, *A*, *B*, *C*, etc. Subject *A* was told he was competing against all the other *A*s in the experiment, subject *B* against all the other *B*s, and so on.) Each person was on his own, trying to maximize his own score. It was emphasized that "there is now no group."

Following the *modal* condition, *S*s were told there would be a brief reversion to the normative situation for a third set of problems, immediately after a third questionnaire had been filled out. This procedure was intended to permit us to reconstitute the group, psychologically, as a frame of reference for asking certain questions. There was no third period, however. Instead, it was announced that the experimental session was over. All deceptions were then explained fully, and care taken that *S*s had time to express their feelings and to receive adequate reassurance and explanation. Before leaving, each *S* made a commitment not to divulge the procedures to others.

RESULTS AND DISCUSSION

Validation of Experimental Conditions

Data were available from the questionnaires by which we could determine how well we had been able to induce the desired conditions.

TABLE 2

FREQUENCY WITH WHICH OTHERS IN A GROUP WERE INCLUDED AS MEMBERS BY THOSE IN THE "HIGH ACCEPTANCE" CONDITION PRIOR TO THE NORMATIVE TASK

Experimental Condition of Recipient	Responses on Questionnaire I	
	Not Included	Included
High acceptance	10	181
Nonacceptance	43	25

Note.—$\chi^2 = 10.36$; $p = <.01$.

TABLE 3

MEAN ATTRACTION TO THE GROUP

Person-group Relationship	Experimental Conditions	*N*	Initial	Medial	Final
Psychological Membership	(high att., high acc.)	33	7.4	8.8	8.4
Marginal Group	(low att., high acc.)	35	7.7	7.8	7.9
Preference Group	(high att., nonacc.)	10	7.8	8.7	8.1
Psychological Nonmembership	(low att., nonacc.)	8	5.6	6.9	6.8

1. *Acceptance as a member.* Our conceptual definition of *acceptance* involved locating the accepted person within the power field of the group; or more simply, recognizing him to be a member of the group. The following question was asked immediately preceding the normative condition:

> Please indicate what contribution you think that *each member of the group* will make to the total group score on the *Group Prize* task you are about to do.

Respondents were asked to rank the members of the group. An appropriate validation of *acceptance* or *nonacceptance* is whether or not an *S* was included as a member of the group, i.e., was assigned any rank whatsoever on this item. In Table 2, the hypothesis of independence can be rejected with a high degree of confidence. Nonaccepted *S*s were included as a "member of the group" significantly less often than accepted *S*s.

Almost the identical question was asked after completion of the modal task condition. The results from this item were similar to those in Table 2, χ^2 being 12.54, significant at $<.001$ level. It appears that the desired difference between *high acceptance* and *nonacceptance* conditions was created and persisted throughout the experiment.

1·9·70
J280

2. *Attraction to the group.* The conceptual definition of *attraction* was "the resultant force acting on an individual to remain in or locomote into a group." The following item was included in all three questionnaires:

> If you were given the opportunity at this point to change to a different group of fellows meeting next door, what would you do?

The possible alternative responses, on a 10-point scale, varied from "I'd very much want to remain a member of this group," to "I'd want very much to change to another group." The mean scores from this item, presented in Table 3, show *S*s in the high attraction condition having greater preference for their own group than those in the low attraction condition. When we combined acceptance conditions, however, only on the second questionnaire was the difference between high attraction and low attraction means significant at the $<.01$ level, using the t test. One possible explanation for our failure to create larger differences in attraction between the two high acceptance conditions is that being accepted in a group may increase a person's attraction to it. Thus, in spite of our attempt to create low attraction for those in the *marginal group relationship*, they do not differ much from those in the *psychological membership* condition. It appears, therefore, that we were successful in creating differences between the two attraction conditions, but not very large differences, measured by this criterion.

3. *Task conditions: normative and modal.* After the completion of both experimental periods, we asked *S*s: "How much do you think that the members of the group wanted to influence your answers?" on each set of problems. In both high and low attraction conditions, *S*s attributed to others more desire to influence them in the *normative* period than in the *modal*. The differences were both significant at the $<.001$ level, using two-tailed t tests. This constitutes indirect evidence of the validity of the two task conditions.

Results for Conformity

The measure of conformity used was the proportion of person-trials on which change to the judgment of the divergent majority occurred. For each *S* on each trial we first considered whether or not his initial judgment (expressed by the notes he had sent to others) was the same as or different from that of the spurious unanimous majority judgment, ignoring the degree of confidence expressed.[6] Only those cases where these judgments differed were included in the analysis. On each trial, an *S*'s final judgment was then categorized as *changed* or *did not change* to the judgment of the majority. The results of this analysis are presented in Table 4.

TABLE 4
PROPORTION OF TRIALS ON WHICH CHANGE OCCURRED TO THE UNANIMOUS DIVERGENT JUDGMENT OF THE MAJORITY

Person-group Relationship	Experimental Conditions	N^a	Normative Task Condition	Modal Task Condition	t
Members	high att., high acc.	33	.622 (a)	.376 (e)	5.12**
Marginals	low att., high acc.	35	.492 (b)	.331 (f)	3.16*
Preferences	high att., nonacc.	10	.564 (c)	.560 (g)	***
Nonmembers	low att., nonacc.	8	.523 (d)	.510 (h)	***

* $p < .01$.
** $p < .001$.
*** p not significant by inspection.
[a] In the Modal condition, both the *Marginals* and *Nonmembers* conditions had one less subject than indicated above.

If we compare these results with the predictive model in Table 1, we find that the data correspond to the predictions made for *S*s in the *high acceptance* condition but not for those in the *nonacceptance* conditions. The first five hypotheses, all concerning the conformity behavior of accepted persons, are supported by the results in Table 4. Hypotheses 6 to 13, however, which make predictions involving the conformity behavior of nonaccepted persons, are not confirmed by our results. First, the implications of the findings regarding accepted persons will be discussed and some additional results presented that are relevant to this discussion. Then we shall undertake to explain our failure to predict the conformity behavior of nonaccepted persons by a reexamination of the assumptions underlying the theory and operational definitions.

Conformity behavior is greater in the normative than in the modal condition for both Members and Marginals, as predicted by the first two hypotheses. Members also conform

---

[6] Our initial analysis was performed using a scale measure of conformity which included the degree of confidence. This yielded results which, although not substantially different from those reported here, were less clear. We are indebted to Leon Festinger for suggesting the later analysis.

TABLE 5
MEAN NUMBER OF NOTES SENT PER POSSIBLE RECIPIENT

Person-group Relationship of Sender	Normative Task Condition		Modal Task Condition	
	Accepted Recipients	Nonaccepted recipients	Accepted Recipients	Nonaccepted recipients
Members	8.4	4.5	8.5	7.9
Marginals	7.9	4.0	7.7	6.3
Preferences	5.9		5.8	
Nonmembers	6.0		6.0	

more than Marginals in the normative condition, providing support for the third hypothesis (*t* is 1.88 for 66 *df*, significant at about the .06 level).[7] It follows that Hypothesis 4 is confirmed: Members in the normative condition conform more than Marginals in the modal (*t* is 3.63 for 66 *df*, significant at the <.001 level of confidence). The fifth hypothesis, which predicted that Marginals in the normative condition would conform more than Members in the modal, is also supported, although with somewhat less confidence, by the results in Table 4 (*t* is 1.63 for 66 *df*, with $p = .10$). Since these five hypotheses were all derived from the two assumptions about *social reality* and *group locomotion* forces to conform, they appear to provide support for these theoretical statements. It seems especially likely that the latter assumption is trustworthy, since it has generated a number of successful predictions and can scarcely be saddled with the responsibility for failing to predict the conformity behavior of nonaccepted persons.

Some additional findings lend support to the group locomotion postulate. If a group of interdependent persons induce forces to conform upon all those accepted as members, but not upon those unaccepted, in areas relevant to *group locomotion*, this pattern of influence attempts should be reflected in the volume and direction of communications that were sent during the experiment. In Table 5 is presented the mean number of notes, adjusted for number of possible recipients, sent in each experimental condition. In the normative condition, both Members and Marginals sent far more notes to accepted than to nonaccepted recipients (*t* is 7.09 for Members and 10.83 for Marginals, with $p < .001$ in each case). Thus when a group of persons are interdependent with respect to a common goal, they apparently consider others whom they accept as members to be more appropriate targets of communication than those whom they do not accept.

When forces to conform to others' inductions are acting upon members of a group, it is likely that these persons will experience some desire to behave in the induced direction. On the final questionnaire, *S*s were asked the following question concerning both the normative and modal task situations:

> During the Group-Prize (Position-Prize) set of problems, how much tendency did you feel to give the same answers as the other fellows?

The mean ratings (on a 10-point scale) for each experimental condition are presented in Table 6. Both Members and Marginals felt a greater tendency in the normative condition than in the modal to give the same answers as the other persons in their group. Such was clearly not the case for persons in the nonaccepted conditions. Thus the accepted persons not only conformed more in the normative condition than in the modal, but they were also aware of a greater desire to do so, presumably because, in the normative condition but not in the modal, they felt the pressure of other members' expectations.

Let us turn now to consider the unconfirmed hypotheses about the conformity behavior of nonaccepted persons. The model in Table 1 calls for conformity behavior of Preferences and Nonmembers to be the same in both normative and modal conditions, and not different from the conformity behavior of Members and Marginals in the modal condition. These predictions derive from the assumption that *social reality* forces to conform, of equal magnitude, are the only such forces

TABLE 6
MEAN RATING OF OWN DESIRE TO GIVE SAME ANSWERS AS OTHERS

Person-group Relationship	Normative Task	Modal Task	*t*
Members	6.7	5.2	3.97**
Marginals	5.9	5.4	2.04*
Preferences	6.8	6.6	
Nonmembers	5.6	5.6	

* $p < .05$.
** $p < .01$.

---

[7] The two-tailed *t* test was used throughout the study, regardless of whether or not the finding was predicted in advance.

TABLE 7
MEAN ESTIMATE OF OWN RELATIVE ABILITY

Person-group Relationship	Mean Rating
Members	5.9
Marginals	6.1
Preferences	4.8
Nonmembers	4.0

acting upon persons in these six cells. An examination of Table 4 suggests, however, that some unpredicted additional force to conform was acting upon Preferences and Nonmembers in both normative and modal conditions. In all four cells their conformity behavior is uniformly higher than was anticipated.

A possible explanation lies in our experimental manipulation to create the *nonacceptance* condition. By giving the nonaccepted *S* information that his ability on the preliminary set of problems was poor, it is likely that his confidence in his ability to perform the subsequent tasks was lowered. Hochbaum (7) found that persons lacking in self-confidence were less able to withstand group pressures to uniformity. Where a person has little confidence in the information provided by his own perception of a stimulus, he will have an enhanced need for *social reality*.

On the final questionnaire *S*s were asked the following:

> Please compare your own ability on the problems to the average ability of the other persons present (excluding the two researchers).

The results of these ratings are presented in Table 7. Persons in the *nonacceptance* conditions clearly estimate their ability to be less than do those in the *high acceptance* conditions (the difference between Members and Preferences is significant at the .07 level, and that between Marginals and Nonmembers at $<.001$ level of confidence). It is probable, therefore, that the unanticipated additional force to conform acting upon nonaccepted *S*s derived from their greater need for social reality created by the experimental induction of nonacceptance. Whether a person who is not accepted as a member of a group for reasons unrelated to his ability to perform a particular task will also have similar forces acting upon him to conform is a question that needs to be answered.

An alternative explanation to the one offered above is suggested by our experiences in running this experiment, and certain tendencies in the data too weak to be discussed here. Our observations led us to believe that nonaccepted persons were experiencing considerable anxiety. They were greatly relieved, almost without exception, when informed in the postexperimental interview about the deception that led to their exclusion from the group. A closer examination of the psychological situation of the person who is not accepted as a member of a desirable group indicates that the experience may be quite disturbing, since the person's ego-processes are under threat. Under such circumstances, the person may have what might be called a *need for social reassurance*, created by his exclusion[8] from the group on invidious grounds. Such a need might lead the person to defend against rejection by identifying with the members of the group and conforming closely to their opinions and judgments.

SUMMARY AND CONCLUSIONS

An experiment was designed to test hypotheses derived from assuming distinctive processes called *social reality* and *group locomotion*, each of which generates forces to conform under specified conditions. Four types of person-group relations were created by experimentally varying subjects' attraction to the group and acceptance as a member. A modification of Asch's line problem and experimental situation was used to test the conformity behavior of subjects.

The results are in accord with hypotheses advanced about the conformity behavior of highly accepted persons, thus supplying evidence in favor of the assumptions about social reality and group locomotion processes. An analysis of communication patterns within the group, and of subjects' feeling about wanting to answer the same as other persons in the group, also support the conclusion that in a goal-oriented group of interdependent persons working upon a relevant task, forces to conform are induced and perceived by all persons accepted as members.

A number of hypotheses, derived from the same theoretical assumptions, were also advanced about the conformity behavior of nonaccepted persons. These hypotheses failed to predict correctly, since the conformity behavior of these subjects was uniformly higher than anticipated. It is likely that the experimental manipulation of nonacceptance created an enhanced need for social reality in excluded subjects: they perceived their own ability on the experimental task to be relatively low. These additional social reality forces to conform may have been responsible for the unpredicted high level of conformity of nonaccepted subjects.

---

[8] A re-examination of our operations in this experiment leads to the conclusion that we created *negative acceptance* rather than *nonacceptance*, as intended, since *S*s were not just located outside the boundary of the group but were barred from membership. This distinction is discussed more fully elsewhere (8).

An alternative explanation is also proposed for the high level of conformity behavior of nonaccepted persons: that they were really *rejected* rather than *not accepted*, and that this created in them a *need for social reassurance*. It is suggested that in response to this need, they had forces acting upon them to regain membership in the group by identifying with its members and conforming to their judgments.

REFERENCES

1. Asch, S. E. Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men: Research in human relations*. Pittsburgh: Carnegie Press, 1951.
2. Asch, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
3. Back, K. W. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–23.
4. Deutsch, M., & Gerard, H. B. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 629–636.
5. Festinger, L. Informal social communication. *Psychol. Rev.*, 1950, **57**, 271–282.
6. Festinger, L., & Thibaut, J. Interpersonal communication in small groups. *J. abnorm. soc. Psychol.*, 1951, **46**, 92–99.
7. Hochbaum, G. M. The relation between group members' self-confidence and their reactions to group pressures to uniformity. *Amer. sociol. Rev.*, 1954, **19**, 678–787.
8. Jackson, J. M. A space for conceptualizing person-group relationships. *Hum. Relat.*, in press.
9. Jackson, J. M., & Saltzstein, H. D. *Group membership and conformity processes*. Ann Arbor, Mich.: Research Center for Group Dynamics, Univer. of Michigan, August 1956.
10. Schachter, S. Deviation, rejection, and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190–207.
11. Sherif, M. A study of some social factors in perception. *Arch. Psychol., N. Y.*, 1935, No. 187.

Received May 9, 1957.

VARIABILITY OF LIGHT PERCEPTION THRESHOLDS IN BRAIN-INJURED CHILDREN[1]

HENRY J. MARK
Johns Hopkins University, Department of Otolaryngology
AND BENJAMIN PASAMANICK
Ohio State University, Department of Psychiatry

THIS paper reports some differences in light perception threshold measurements found between brain-injured (pyramidal tract damage) and control children. The results were obtained in the course of a larger study aimed at discovering neuropsychological measurements which might be used to detect otherwise unnoticed alterations in cerebral functioning. Attention was directed at intra-individual threshold variability, rather than the average threshold level alone, and the brain-injured children were expected to have more "diffuse" or variable thresholds than the normal controls. Increased intra-individual threshold variability was hypothesized as a "fundamental disturbance" in Goldstein's sense of that term (2), i.e., it was considered that the effect of a localized lesion may be detectable in many cortical areas, thus directly or indirectly involving not only the primary sensory and motor projection areas but also those areas assumed to mediate more complex cortical functions.

Intra-individual response variability has been of general interest in psychology in such different areas as test construction and psychophysical threshold measurements. In clinical psychology, intra-individual variability *among* the subtests of intelligence scales is of diagnostic interest and is often referred to as "scatter of function," representing an *S*'s ability level in different intellectual areas of functioning. In clinical audiology, interest now is focused on repeated threshold determinations, increased threshold variability *within* the auditory sense modality being considered an important factor in distinguishing central hearing or language disorders from end-organ impairment.[2] This report restricts itself to the results obtained in light perception thresholds. Studies with repeated critical flicker fusion threshold measurements (3) and other visual phenomena (4), as well as with threshold measurements in other sense modalities (5) in the same *S*s as those used here, report similar findings of increased intra-individual variability in the brain-injured. One may therefore hypothesize that organic dysfunction readily manifests itself in lack of consistency of one sort or another; and, indeed, it is easy to conceptualize how increased variability in so-called simpler primary functions may actually give rise to dysfunction in processes such as conditioning and memory as well as more complex behavior patterns.

In addition to *variability* of threshold, comparison between the brain-injured *S*s and controls in the *magnitude* of the thresholds (i.e., exposure time required) is of interest in view of the findings in the CFF experiments with the same groups that only under a few conditions were there lowered CFF thresholds in the brain-injured. It was hoped that the effects of a presumably distant cortical lesion on visual efficiency would be more readily detectable in the light perception thresholds than in the CFF thresholds of the brain-injured children tested.

The threshold point is conventionally taken as that stimulus magnitude for which the probability of a response ("light seen") is 50 per cent. In these experiments, the estimated threshold is always the mean of ten threshold measurements. By hypothesis, higher thresholds (longer exposure time), corresponding to loss of visual efficiency, should be associated with the brain-injured. The estimated spread or variability is reported as the *V* score, and is always taken to be the logarithm[3] of the variance of the distribution of measurements around the mean, i.e., around the thresholds. Accordingly, brain-injured *S*s should show larger *V* scores (greater intra-individual variability).

---

[1] This research was supported by an Alfred P. Sloan Foundation, Inc., grant to the Division of Laryngology and Otology, Department of Surgery, Johns Hopkins University, as well as by a grant from the Foundation for Mentally Retarded and Handicapped Children of Baltimore. The authors wish to thank Paul Meier for invaluable assistance and advice, and Victor Laties and Bernard Weiss for their critical reading of the manuscript.

[2] Hardy, W. G. Personal Communication, May 1957.

Measurement of spread is complicated by the fact that the observed S-shaped psychometric curve consists of what is traditionally called a sensory or "true" component plus a measurement error. If this measurement error is large relative to the difference in spreads between two *S*s, then the resultant large spread due to measurement error will obscure any difference in spread between two *S*s which may in fact exist. An attempt was therefore made to refine apparatus and technique in order to reduce inherent measurement error to a magnitude where it would not interfere with distinguishing abnormal spreads.

METHOD

Subjects

The brain-injured group was composed of ten children (six males, four females) with diagnosed pyramidal tract involvement and with no primary diagnosed ophthalmologic pathology. The control group was composed of ten children (five males, five females) with diagnosed conditions which did not justify an assumption of brain damage or injury, such as Erb's palsy, osteochondritis, or tuberculosis of the spine. All *S*s attended a Baltimore school for handicapped children. The groups were roughly comparable in terms of age, IQ, and sex, and the effects of these covariates on all significant discriminators were examined by means of regression diagrams. None was significantly related to performance. For example, there was no apparent regression of thresholds or *V* scores on IQ (which ranged between 64 and 121), and it was therefore assumed that IQ did not enter as an important experimental factor in the difference between brain-injured and control *S*s. The mean IQ of the brain-injured group was 89 with an *SD* of 16; the mean for the control group was 96 with an *SD* of 16. The age range was between 10 and 15 years with a mean of 12 for both groups.

Apparatus

The apparatus consisted of two Sylvania R 1131 C glow-modulator tubes mounted centrally on a perimeter so that two focused lines of light, each ¾ mm wide and 6 mm long and separated by a distance of 4 mm could be projected by means of a simple achromatic lens onto a piece of white plastic mounted on black velvet cloth. The glow modulator was activated by an electronic pulse-type stimulator and timer (6) which permitted continuous and independent variation of exposure times, t_1 and t_2, of the two light stimuli, as well as continuous and independent variation of the pause (p) between the onset of the two light stimuli. The two rectangular focused lines of light reflected by the white plastic were viewed by the *S*s from a distance of 66 centimeters. In both parts of this experiment, the exposure time of the left light (t_1) was the only experimental variable and is reported in milliseconds (ms). The exposure time of the right light (t_2) was held constant at 135 ms throughout. In Part I (absolute light thresholds), the pause between the onset of the two lights was zero (i.e., they were presented simultaneously); in Part II (apparent movement thresholds), the pause between onset of the two lights was held constant at 150 ms, which, under these conditions, invariably gave rise to an apparent movement phenomenon. Perception of such movement and reported cessation of movement was thus considered evidence of perception of the left stimulus light. The frequency of presentation of the two lights was electronically timed at one per second throughout both parts. The experiments were carried on in total darkness. With constant exposure, the reflected lights had a brightness of the order of one-tenth millilamberts (0.1 ml) as measured with a MacBeth Illuminometer. (In order to have an adequate range for threshold measurements of t_1, the brightness of the left light was actually somewhat less than that.)

Procedure

For Part I (absolute light thresholds), two stimulus lights were presented simultaneously once per second. The exposure time (t_1) of the dimmer (slightly reddish) light was varied continuously from .2 ms to 9.3 ms. The first measurement was the descending threshold (L_{Desc}) where t_1 was continuously decreased. *S* reported when he was no longer able to see the left light (but of course continued to see the right light). Then an ascending measurement (L_{Asc}) was obtained where t_1 was continuously increased. *S* reported when he again saw the left light next to the right light. This procedure was repeated ten times. In this way two thresholds and two *V* scores (L_{Desc} and L_{Asc}) were obtained for every *S*, with each threshold based on 10 measurements and each *V* score therefore associated with 9 degrees of freedom. Before these measurements were made, the *S*s were given practice in making these reports.

---

[3] The *V* scores are given as $4 + \log_{10} s^2$ (rather than s^2 or s) for the following reason. If we suppose that the psychometric function can be described as a cumulated normal distribution, we know that the distribution of s^2 is that of a constant *times* a mean square (chi square/degrees of freedom) distribution, i.e., s^2 is distributed as $\sigma^2(\chi^2/df)$, where σ^2 is the variance of the psychometric curve. Consequently, the distribution of $V = \log s^2$ is that of a constant *plus* the logarithm of a mean square, i.e., $V = \log s^2$ is distributed as $\log \sigma^2 + \log (\chi^2/df)$. Since σ^2 is a parameter of the curve, the precision of *V* depends only on the (known) distribution of the second term. Thus, since 10 measurements were made for each threshold, the *V* scores all have equal precision, and the variance of a single *V* score is 0.0469 (1). An approximate test for equality of a set of *k* variances is given by calculating the ratio of $\Sigma(V_i - \overline{V})^2/(k - 1)$ to 0.0469 and comparing the result with the appropriate percentage point of the $F_{(k-1,\infty)}$ distribution.

It must be emphasized that it is the *heterogeneity* rather than the absolute magnitude of the individual σ^2 values which affects the value of $\Sigma(V_i - \overline{V})^2/(k - 1)$. If individual differences in spread are large compared with measurement error, the mean square will be large. However, if measurement error is *increased* so that the differences in spread are obscured, the mean square will tend to be *reduced* and the differences in spread may not be detected.

The same procedure was followed for Part II (apparent movement thresholds), but the lights were presented sequentially with a pause of 150 ms between the onset times of the two lights. (A pilot investigation with children outside the study indicated that such a pause always produced a movement effect under these physical conditions.) Continuous decrease of t_1 had no other obvious effect than to stop the "movement," leaving the line blinking on and off in the same place once per second. An increase of t_1 again produced the movement effect. In this part of the experiment it was more convenient to have an ascending measurement (M_{Asc}) always precede a descending measurement (M_{Desc}), signifying the beginning and cessation of the movement effect respectively. The two thresholds obtained were again based on 10 measurements each, and the corresponding *V* scores are therefore again associated with 9 degrees of freedom.

The instructions, aimed at minimizing differences between ascending and descending thresholds follow:

For L_{Desc}: You now see two lines,[4] one brighter than the other. You will notice that the bright light will always be there, and will not change. The dim light is a little red and is the one to watch carefully. The dim light will get dimmer and dimmer so that you can hardly see it any more until it is all gone. Let me know as soon as you notice the dim light is not there any more. Let me know as soon as you notice only the bright line without the dim light next to it.

For L_{Asc}: Now you see only the bright line. Let me know as soon as you notice that there is a dim light next to it again. As soon as you notice that the bright line is not just blinking on and off alone, but that there's a dim light next to it again.

For M_{Asc}: Now you see one line that is standing still and just going on and off. Let me know as soon as you notice that the line is moving across a little bit—as soon as you notice it is *not* standing still but moving across from left to right even a little bit.

For M_{Desc}: Now you see one line that is moving across from left to right. Let me know as soon as you see that it's *not* moving across any more, that it is standing still; that it is just blinking on and off, but not moving across any more.

Nonparametric rank methods (7) were indicated in view of the sometimes erratic estimates of spread (*V* scores) of the psychometric functions. The .05 level of significance was adopted as the criterion for rejecting the two-tailed null hypothesis of no difference between groups.

RESULTS AND DISCUSSION

TABLE 1

AVERAGE LIGHT PERCEPTION THRESHOLDS (L_{Desc} AND L_{Asc}) AND APPARENT MOVEMENT THRESHOLDS (M_{Desc} AND M_{Asc}) IN MILLISECONDS FOR THE BRAIN-INJURED AND CONTROL GROUPS OF TEN *S*s EACH

Thresholds	Brain-Injured $\bar{X}$	Brain-Injured *SD*	Control $\bar{X}$	Control *SD*	Diff.
Absolute Light Perception					
L_{Desc}	2.9	1.0	1.9	.5	1.0*
L_{Asc}	2.8	.8	1.8	.5	1.0**
Apparent Movement Perception					
M_{Desc}	1.8	.5	1.3	.5	.5*
M_{Asc}	2.4	1.0	1.6	.2	.8*

* Significant at the .05 level of confidence.

** Significant at the .01 level of confidence.

TABLE 2

AVERAGE LIGHT PERCEPTION (L_{Desc} AND L_{Asc}) AND APPARENT MOVEMENT (M_{Desc} AND M_{Asc}) *V* SCORES (LOG VARIANCE SCORES +4) FOR THE BRAIN-INJURED AND CONTROL GROUPS OF TEN *S*s EACH

Thresholds	Brain-Injured $\bar{X}$	Brain-Injured F^a	Controls $\bar{X}$	Controls F^a
Absolute Light Perception				
L_{Desc}	3.60***	1.80	2.80	.91
L_{Asc}	3.60***	5.11	2.81	3.08
Apparent Movement Perception				
M_{Desc}	3.77***	2.83	3.11	1.33
M_{Asc}	3.55***	6.53	2.75	3.75

*** Significantly different from control group at the .001 level of confidence.

[a] $F_{9,\infty}$ = variance of *V* scores/.0469. If in fact all *S*s within a group had the same spread, then an *F* ratio of 1.88 would occur only 1 time in 20. See text and Footnote 3.

Threshold results are shown in Table 1. Higher average thresholds (longer exposure time) were found in the brain-injured than in the control *S*s. This supports the hypothesis that visual efficiency is impaired in the brain-injured. The difference between the groups is significant both for the absolute light perception thresholds ($p < .02$) and for apparent movement thresholds ($p < .05$). Since CFF thresholds did not discriminate as sharply between these groups (3), light perception thresholds may be more sensitive to a presumably nonoccipital lesion than CFF thresholds.

The average *V* scores of the brain-injured group were significantly larger than those of the control group (Table 2). This confirms the major hypothesis that the brain-injured show increased intra-individual variability. The difference between the groups is significant for the ascending and descending absolute light perception thresholds as well as for the ascending and descending apparent movement thresholds ($p \leq .001$).

---

[4] At near threshold values, the dim line did not look so much like a line but just more like a thin strip of light. In order to obtain the lowest possible thresholds, *S*s were trained to respond to any light in addition to the bright line. This actually was only an issue during the training period with three *S*s and was readily clarified.

The large difference in average V scores between the two groups suggests that this measure may be useful for diagnosing brain injury. In addition, for those tasks in which significant F ratios were obtained (as in the ascending measures of both groups), the technique may be sensitive enough to reflect individual differences within a group. (Significant F ratios, as described in Footnote 3, indicate that the interindividual variability within a group is greater than would be expected by chance.) Thus, from the viewpoint of diagnostic potential, the ascending V scores suggest the possibility that individuals within dichotomized groups may be ranked on a variability scale, which may then be standardized and correlated with other measures or clinical categories.

In a larger study (5), further experiments were conducted to determine whether differences found in the above groups could be reproduced with other groups. By lowering the age limit to eight years and making some changes in procedure, seven additional brain-injured *S*s could be tested.[5] Significant differences in thresholds and V scores between this younger group of seven brain-injured *S*s and seven comparable control *S*s were found, thus supporting the findings with the older groups reported here.

It is also of interest to note that the light exposure time necessary to produce the apparent movement effect was considerably shorter (an average of .6 ms) than the exposure time necessary for the absolute light perception thresholds in 17 out of 20 *S*s tested. It may be that under these experimental conditions, the apparent movement phenomenon is a more sensitive index to sensory excitation than the absolute light perception threshold.

SUMMARY

Magnitude and intra-individual variability of absolute light perception and apparent movement thresholds of ten children with pyramidal tract damage were compared with the threshold and variability of ten non-brain-injured handicapped children of comparable age, IQ, and sex distribution. The results show that the brain-injured *S*s displayed significantly higher thresholds than the control *S*s, thus supporting the hypothesis that visual efficiency may be lowered demonstrably in brain-injured *S*s with presumably nonoccipital lesions. Moreover, as predicted, the brain-injured showed consistently greater intra-individual variability than the control *S*s. When viewed in light of similar studies, the results support the major hypothesis that organic dysfunction gives rise to lack of consistency of one sort or another. The diagnostic potentialities of the employed experimental procedures were discussed.

REFERENCES

1. BARTLETT, M. S., & KENDALL, D. G. The statistical analysis of variance-heterogeneity and the logarithmic transformation. *J. Roy. statist. Soc.* (*Suppl.*), 1946, **8**, 128–138.
2. GOLDSTEIN, K. *The organism.* New York: American Book, 1939.
3. MARK, H. J., MEIER, P., & PASAMANICK, B. Variability of critical flicker fusion thresholds in brain-injured children. *Arch. Neurol. Psychiat.*, in press.
4. MARK, H. J., & PASAMANICK, B. Asynchronism and apparent movement thresholds in brain-injured children. *J. consult. Psychol.*, 1958, **22**, 173–177.
5. MARK, H. J. Studies of perception in brain-injured children. Unpublished doctoral dissertation, Johns Hopkins Univer. Sch. of Hygiene and Publ. Hlth., 1955.
6. ROUSH, R. G., & URBANSKI, E. T. Universal medical timer and pulse stimulator. *Electronics*, 1953, **26**, Part II, 154–157.
7. WILCOXON, F. *Some rapid approximate statistical procedures.* New York: American Cyanamid, 1949.

Received June 10, 1957.

---

[5] In order to keep the groups comparable in terms of experiential background, educational experience, and immediate testing environment, children from other schools or outside the school system were not included in the study. One of the authors (Pasamanick) was consultant for handicapped children with the Baltimore school system. His experience suggests that except for two or three children in another school, Negro before desegregation, and a very small number of children, usually with lesser involvement, scattered throughout the city, the total of 19 brain-injured children tested throughout the year in the larger study constitutes almost the entire population of pyramidal tract damaged children of school age satisfying the conditions described and excluding those meaningfully placed into the "mental defective" category.

THE EFFECT OF SUBLIMINAL STIMULATION UPON AUTONOMIC AND VERBAL BEHAVIOR

N. F. DIXON

University College, London

OVER the last fifty years a number of researches (2, 8, 15, 18) have shown that subliminal[1] stimuli may exercise a selective function upon verbal behavior. A common feature of these experiments was that *S* made his guesses from a strictly limited set of possible responses. In this situation, the subliminal stimulus, whether auditory or visual, was found to increase the probability of *S*'s responding verbally in terms of what was currently presented.

The results from an earlier research (6) by the writer, however, have suggested that subliminal stimulation may also prove influential in a free choice situation. Asked to say "the first word that comes to mind" during subliminal auditory stimulation with words, *Ss* showed a significant tendency to produce autistic associations to the stimuli. Moreover, a significant correlation was found between response latencies and the emotional value of the stimulus items. The first of these phenomena was noted as being consistent with the informal observations of Bruner, McGinnies, and Licklider[2] that prerecognition hypotheses, in tachistoscopic experiments and word articulation tests, were often associations of an autistic kind to the stimulus material. The finding is also consistent with the experimental conclusions of Fisher (9) and Poetzl (14).

The present pilot study, some aspects of which have been covered in a previous paper (7), is concerned with investigating the possibility of these phenomena when the stimulus is visual. It is designed, moreover, to test the subception hypothesis as put forward by Lazarus and McCleary (14). This hypothesis, which postulates a heirarchy of response thresholds, derives from the finding that GSRs conditioned to verbal stimuli could still be evoked below the verbal recognition threshold. Based on this finding, however, the subception hypothesis has been the object of considerable dispute.

In particular, Howes (11) and more recently Eriksen (9) have questioned it on the grounds that the subception effect could be an artifact arising from the Lazarus and McCleary experimental design. The essence of their argument is that a comparison of GSRs and verbal responses can only be meaningful provided *S* has an equal opportunity of demonstrating his discriminatory capacity in the two response systems. In the Lazarus and McCleary experiment this condition clearly did not obtain. With an infinite number of possible GSR categories and only ten verbal response categories, *S* could signal partial discrimination by the one response system but not by the other or, to quote Eriksen, ". . . if the number of verbal response categories is too few we run the risk of spuriously reducing the correlation between the verbal response and the stimuli, and of increasing the partial correlation between the stimuli and the GSR." In support of this argument, it may be added that at least two researches (3, 5) have shown that partial discrimination of briefly exposed verbal material can in fact occur and may be measured by applying suitable operations and controls.

The following experiment attempts to circumvent these criticisms in two ways; first, by allowing *S* a potentially infinite range of verbal response categories and, second, by ensuring total unawareness of the stimulus presentation—a condition not necessitated by tachistoscopic exposure below the recognition threshold.

Hypothesis 1. That the visual presentation of words at an intensity below the awareness threshold would influence the selection of verbal responses in a free choice situation.

Hypothesis 2. That this selective function of the subliminal stimulus would result in verbal associations to the stimulus items.

---

[1] For the purpose of this paper, a "subliminal" stimulus is one of whose nature and presence *S* remains wholly unaware—i.e., its intensity is below the awareness threshold.

[2] Personal communications.

Hypothesis 3. That subjects presented subsequently with their responses would be able to say to which stimulus word each of these had been given.

Hypothesis 4. That galvanic skin responses and verbal response latencies recorded during stimulation with emotionally charged words would exceed those recorded for neutral items.

METHOD

Subjects

Seven *S*s took part, of whom five were male. All the *S*s were undergraduates, four being students of psychology.

Apparatus and Design

The design of the experiment involved four stages:

1. Determination of the difference threshold for brightness.
2. Subliminal stimulation at a controlled intensity and for particular periods of time, with presentation at regular intervals of a supraliminal signal to respond. During this stage, verbal responses, GSRs, and the response latencies were recorded.
3. A word association test in which immediate associations were obtained to the verbal responses given at Stage 2.
4. A matching test in which *S*s were required to say to which of the stimulus items each of their responses had been given.

For threshold determination and subsequent subliminal stimulation, the stimulus material, on 35-mm film, was projected upon an opal flashed translucent screen, image brightness being controlled by a pair of opposed neutral variable filters with a continuous density gradient of .1 per inch. Exposure of the stimulus material was by means of a shutter. The response signal was given by the momentary appearance, in the upper half of the screen, of a small spot of light. This signal was timed to occur two seconds after the shutter opened.

S's verbal responses closed the shutter after each exposure by operating a voice key. Reaction times between the appearance of the signal light and the verbal responses were recorded on a chronotron. GSRs following each exposure were measured on a Pye Scalamp galvanometer in a conventional bridge circuit, electrical contact with the skin being mediated by a pair of nonpolarizing silver silver-chloride electrodes each immersed in a finger beaker containing saline solution.

The stimulus material consisted of 12 items, 10 words and 2 straight lines (one being vertical, the other horizontal). Six of the words were of primary emotional interest, the other four being 'emotionally neutral.' The items were presented in the following order: *this*, [*horizontal line*], *penis*, *barn*, *father*, [*vertical line*], *vagina*, *seven*, *male sex organ*, *line*, *mother*, *female sex organ*.

Procedure

The *S* was seated in a dark room one meter from and facing the translucent screen illuminated to present a uniform brightness of .3 log ft lamberts. A horizontal line of light, subtending a visual angle of 2°48′, was projected onto the center of the screen at such an intensity as to be clearly visible. *S* was instructed:

> I am going to make the line gradually dimmer. All you have to do is to watch it carefully and say "stop" the moment it has completely disappeared. Don't say "stop" until you are quite certain you cannot see anything at all upon the screen.

The brightness of the image was reduced at the rate of .1 log ft lamberts per second until such time as *S* gave his verdict. This procedure was repeated with the line in both vertical and horizontal planes until the lowest threshold value had been found, the criterion for "lowest value" being one not equalled or bettered in three subsequent trials.

The first and third fingers of *S*'s right hand were then immersed respectively in the two electrode beakers, the arm being supported on a padded rest with the hand lightly clamped. *S*'s basic skin resistance was balanced against the bridge circuit and measured by substitution.

The following instructions were then given:

> I am going to project words on the screen but they will be too dim for you to see. With each projection however I want you to make a guess regarding the identity of the word shown. Your signal that a word is there and that you should respond will be a small spot of light appearing momentarily at the top of the screen. During the experiment keep watching the screen though you can see nothing on it. Try and maintain a relaxed passive attitude just letting each guess be the first word that comes to mind after the signal light appears.

A practice trial of ten responses was given, with the shutter closed.

This variable wedge was then set to give an image brightness .3 log ft lamberts below the determined threshold and the first stimulus item exposed. Two seconds later, the signal light was presented and the chronotron started. *S*'s verbal response by activating the voice key stopped the chronotron and closed the shutter. The verbal response and response latency were recorded by *E*. A record was also taken of this observed maximum deflection of the GSR galvanometer (as indicated by a hand follower). Five seconds after the skin resistance had returned to the baseline, the next stimulus item was presented and the procedure repeated as before.

After each response, *S* was asked if he had seen anything. (*In no case* did any of the *S*s report awareness of the stimulus.) The entire series of 12 items was presented, in all, four times to each *S*, the order of presentation being the same each time, with each trial following immediately upon the preceding one (i.e., *S* was given no clue to the fact that a series was being repeated). Immediately following this part of the experiment, *S* was asked for his first association to each of the responses he had given.

For the third stage in the experiment, carried out a week later, *S* was given a card upon which were printed in a random order the 12 items of stimulus material that had been presented subliminally. *S* was asked whether the list seemed in any way familiar. In every case the reply was to the contrary. The following instructions were given:

> I am going to read you a list of words [*S*'s original responses randomly ordered]. After each I want you to tell me the number of that item on your list which in your opinion is most appropriate to the word I have given. I shall not ask you for the reasons for your choice which should be entirely personal and not bound by logical or commonsense considerations; just choose the item which you feel, for any reason whatever, is most suited to the word I have given. Take as much time as you like, then give me the number of your choice.

This entire procedure was carried out for all seven *S*s.

RESULTS

Response Latencies

The first hypothesis tested was that response latencies recorded after presentation of emotional items would exceed those recorded after neutral items. An analysis of the data revealed that five out of the seven *S*s who took part showed shorter average latencies for the neutral items than they did for the emotional items. The differences were not statistically significant, however, by the sign test.

TABLE 1
GALVANIC SKIN RESPONSES

Subject	Mean GSRs		Means of Maximum GSRs	
	"Emotional" Items	"Neutral" Items	"Emotional" Items	"Neutral" Items
MM	8.40	8.18	14.87	12.95
MT	15.35	14.05	23.98	21.22
MTp	2.73	1.88	5.28	3.90
MDp	2.52	1.99	4.78	3.68
FB	1.64	0.59	4.38	1.28
FSp	3.87	2.11	7.93	4.00
MBp	5.06	3.21	7.92	5.08

TABLE 2
SIGNIFICANCE OF CORRECT MATCHES

Subject[a]	Variance	Mean Chance Expectation	Observed Correct	p
MM	3.69	4	6	.159
MT	3.59	4	7	.061
MTp	3.61	4	9	.005
MDp	3.55	4	3	.308
FB	2.87	3.5	7	.020
FSp	3.58	4	3	.309
MBp	2.72	3	4	.274

[a] The prefix *M* or *F* denotes the sex of *S*; the suffix *p* indicates that *S* was a student of psychology.

Galvanic Skin Responses

The hypothesis tested was that GSRs accompanying the presentation of emotional items would exceed those recorded for neutral items. The results showed that all seven *S*s gave higher average GSRs for emotional items than they did for neutral ones (see Table 1). By the sign test this result in the direction predicted is significant ($p = .008$).

Verbal Responses

The first hypothesis tested was that when presented subsequently with their responses, *S*s would be able to say to which stimulus word each had been given. For the purpose of this test, a statistical method developed by Stevens (17) was adopted, the advantage of this test being that it takes into account the total variance of all responses made.

From Table 2 it can be seen that of the seven *S*s who participated in this experiment, five matched their responses against the correct stimulus item more often than would be predicted on a chance basis, two of them by an amount that is statistically significant. The

TABLE 3
RESPONSES CORRECTLY MATCHED TO STIMULUS WORDS

Stimulus Word[a]	Original Response by Each Subject						
	MM	MT	MTp	MDp	FB	FSp	MBp
Penis			Run, shower				
Male Sex Organ	Brush						Cheroot
Female Sex Organ						Supply	Follow
Vagina		Shade			Mouth		Focus, triumph
Father		Bulbous, ear			Face, swim, food		
Mother	Screen, pleasant		Silver, sugar		Spot, beat	Species	
Barn	Cartwheel	Cradle	Star, open, wood	Pony		Purple	
Line	Valparaiso	Sidelong, do	Circle	Flight	Light		
Seven	Depth		Crowd	Book			
This		Face					

[a] No correct matches occurred for responses to the vertical or the horizontal line.

TABLE 4
RESPONSES MATCHED TO SYNONYMOUS STIMULUS WORDS

Original Stimulus Word	Synonymous Match	Responses
Penis	M.S.O.	Million, Moore, deny, chair
M.S.O.	Penis	thick, eject, figure, line, red chalk
Vagina	F.S.O.	nostril, ready, glorify
F.S.O.	Vagina	bloodshot, noon

other two *S*s scored slightly below the chance expectation.

It is a convenient feature of the statistical method used that both the deviations from chance and the variances are additive over a number of *S*s, thus providing for an over-all test of significance. By this method, a standard normal deviate of 2.57 is obtained, significant (p .005) on a one-tail test. Table 3 shows the actual responses that each *S* succeeded in matching correctly.

The Matching of Responses to Synonymous Stimulus Words

It will be recalled that the stimulus material included two pairs of synonymous stimulus items: "Penis" and "Male Sex Organ," and "Vagina" and "Female Sex Organ." These were included in order to test the hypothesis that the response to the subliminal stimulus material is a response to the meaning which the latter connotes, the prediction being that there would be a significant tendency for *S*s to match their responses against the synonym of the item which had originally evoked it. These synonymously matched responses are shown in Table 4. Applying the Stevens test to these matches yields a significant value for Z of 2.408 ($p = .008$).

An interesting *post hoc* observation was that no correct matches occurred for the 54 responses given during subliminal stimulation with the vertical and horizontal lines. This effect was not predicted though it does perhaps fit in with a conclusion drawn by the writer from an earlier research that the subliminal stimulus influences verbal behavior by the facilitation of long standing associations between objects and the verbal responses which have accompanied these objects in the past of the individual concerned.

The Nature of the Verbal Responses

Having justified the assumption of a causal relationship between subliminal stimulus and verbal response by the preceding analyses of GSRs and response matches, the extent to which these matches support the hypothesis that the subliminal stimulus evokes verbal associations was further examined in the light of the actual responses given to each of the stimulus items.

The results from this study which are referred to in an earlier paper (7) revealed, first, that many of the responses were likely associations to the stimulus words and, second, that there was a tendency to guess words that were symbolic (in the Freudian sense) of the objects connoted by the stimulus word. These qualitative interpretations were supported to some extent by the *S*'s matchings and to some extent by the associations given to the responses. Thus, to the subliminal stimulus word "Line" the Subject MM responded "Valparaiso." His subsequent association to "Valparaiso" was "Liner."

Introspective Data

For investigations into the effect upon behavior of subliminal stimulation, introspective data are obviously of prime importance in establishing whether or not the stimulation was in fact subliminal. In this experiment, three appeals were made to introspection for evidence on this point: (*a*) in the original threshold determination; (*b*) in a general instruction prior to subliminal stimulation (the instruction was to report immediately during the course of the experiment any awareness of anything on the screen other than the periodic signal light); and (*c*) at the end of the experiment when each *S* was asked whether at any time he had seen anything at all upon the screen. The question of threshold determination has already been dealt with. Regarding the possible reporting of awareness during the experiment, none of the *S*s reported seeing anything. Finally, when questioned after the experiment, the *S*s were unanimous in reporting that they had seen nothing and that their responses had in fact all been pure guesses.

Such other introspective data as the *S*s were able to provide are shown below:

MM	Words just came up when I saw the signal. I had no imagery.
MT	I had visual images of wheat, faces, and birds. [These objects figured amongst his responses.]
MTp	I had many extraneous thoughts. "River" came up twice. I thought of words in an encyclopedia, also of words already given.
FB	Nothing to report.
FSp	Slightly fatigued at end. Apprehensive at beginning about what I might say. Frequently thought of responses amongst which five tabu words came up.
MBp	Nothing else to report.

These introspections reveal little beyond the fact that the *S*s seemed to have little insight into the origin of their responses. The comment that "words just came up" suggests a somewhat passive arrival in consciousness of the response to make, rather than an active process of guessing or selecting, the appropriate response.

DISCUSSION

The subception hypothesis which postulates a heirarchy of response thresholds has met with resistance on two counts. On the one hand it is regarded as uneconomical, on the other as inviting a homuncular theory of perception, particularly when related to the concept of perceptual defense. To avoid what Allport (1) has called the anthropomorphic notion of some inner perceiving agent, other explanations have been given of the relevant phenomena as, for example, "voluntary suppression" in the case of apparent perceptual defense and "a statistical artifact" in the case of the Lazarus and McLeary findings.

The results from the present experiment can not, however, be so easily explained by these means. At the time of threshold determination, the *S*s had nothing to suppress and thenceforth were being stimulated well below this threshold. In the second place, explanation of the differential GSR behavior in terms of partial discrimination at a conscious verbal level seems unlikely for two reasons. In the first place, although *S* had a virtually infinite range of verbal response categories his responses bore no structural similarity whatever to the stimulus items. In the second place, the introspective data confirmed that the *S*s were completely unaware of the stimulus presentations.

In this connection it should be recalled that this experiment differed from those of Lazarus and McCleary on the one hand, and Chapanis and Bricker on the other, inasmuch as stimulation was applied not merely below the recognition threshold but below what for convenience' sake we may describe as the absolute awareness threshold. *S*s knew that they were being stimulated only because they were told so by *E*, *not* because they were aware of something in the visual field.

This raises a problem crucial to the whole concept of subception. How, it may be asked, can we assume that there was no discrimination at a verbal level while at the same time asserting that the verbal responses were associations to the stimulus items? In the writer's opinion, the apparent paradox stems from a confusion that has permeated much recent discussion of the topic. It is the confusion between three quite distinct orders of response—awareness of the stimulus, verbalization about this awareness, and verbal behavior determined by, but subjectively unrelated to, the stimulus. The confusion arises no doubt from the mistaken parallel that is sometimes drawn between repression and certain perceptual processes. Because repression is sometimes regarded as the inability to verbalize ideas and feelings it does not necessarily follow that the ability to verbalize can be taken as a criterion of perceptual awareness.

If ability to verbalize could be so used, then, as one psychologist (13) has recently remarked, young children and aphasics might well be useful subjects for the demonstration of subception merely because they would in all probability react emotionally to stimuli about which they could not verbalize. The suggestion here is that not only are awareness and verbal thresholds far from synonymous but that the relationship between them is considerably more complex than some writers would have us suppose. Consider, for example, some of the possible relations that may be observed:

1. *S* is wholely unaware of the stimulus and verbalizes, "I can see nothing, there is nothing there." No correlation can be found between his other verbal responses and the stimulus.

2. *S* is unaware of the nature of the stimulus but aware of the fact that he is being stimulated. He can detect something in the visual field. He verbalizes, "There is something there but I cannot say what." A partial correlation, however, is found between the structure of his verbal responses and that of the stimulus items.

3. *S* is wholely unaware of the stimulus. He verbalizes, "I can see nothing, there is nothing there," but some correlation is found, not between the structure, but between the meaning area of the stimulus material and that of his responses.

This latter case is exemplified by our present data. In other words, resolution of the apparent paradox depends upon distinguishing between the verbal response system underlying intended verbal report and activated by stimuli above the awareness threshold and that of autistic verbal behavior activated by stimuli below this threshold. Two criteria may be applied to this distinction: first, whether or not *S* reports any awareness of the stimulus and, second, whether the observed relationship between the stimuli and responses is *such as he was programmed* to make. In the present case, the level of intended verbal report was clearly not active. *S* reported no awareness of the stimulus, and the relationship between his responses and the stimuli was not that which he had been instructed to observe.

The evidence then suggests that with this admittedly small sample, partial discrimination did occur below the level of awareness. The data do not require us to regard the threshold for autonomic descrimination as necessarily lower than that for determination of verbal behavior but merely to assume that both these thresholds lie below those of stimulus awareness and intended verbal report. If we allow that stimulus awareness and intended verbal report belong to one response system, while autonomic response or the determination of autistic verbal behavior belong to another, then our data suggest that subception occurred. Whether the autonomic response precedes, is coincident with, or follows upon the verbal behavior remains a matter for conjecture. Since it was the stimulus items rather than the verbal responses that were emotionally laden, the first of these alternatives seems the more likely.

However, while the data support the subception hypothesis, they do not permit us to assume that the differential GSRs were a function of an unconscious emotional disturbance. They could be due to the greater difficulty of associating to a tabu word, as recently demonstrated by Jacobs (11). To what extent the difficulty of associating is a function of the emotional nature as opposed to the mere unfamiliarity of these words remains an open question. That emotional value *is* a factor however is supported by the present finding that the GSRs recorded for "Father" and "Mother" exceeded those recorded for *any* of the neutral words though these were of the same frequency rating. Also in favor of the hypothesis that the higher GSRs recorded for the emotional items do in fact reflect the emotional impact of these words is the decline in GSR with successive presentations. The occurrence of associations often of an autistic kind in the context of a subliminal experiment has been dealt with extensively elsewhere (7). Suffice it to say that it suggests the operation of an ontogenetically determined hierarchy of classificatory levels wherein the subliminal stimulus exercises a selective function upon responses hitherto associated with the stimulus material, the occurrence probabilities of particular associations being determined by emotional as opposed to logical factors; that is, the subliminal stimulus activates *primary* thought processes.

That there is, even in the waking state, a general tendency towards autism in cases of reduced sensory input quite irrespective of particular unconscious discrimination is incidentally suggested by the recent experiments of Bexton, Heron, and Scott (4).

SUMMARY

This pilot experiment was undertaken to examine the general hypothesis that the effect upon verbal and autonomic behavior of subliminal *visual* stimulation would be comparable to that found in the case of subliminal stimulation that was *auditory*. In particular, the design was orientated towards investigating the subception hypothesis of Lazarus and McCleary. The hypotheses tested were that:

1. Response latencies and GSRs would be determined by the affective value of the stimulus material.

2. The verbal guesses made during subliminal stimulation would be associations to the stimulus items as shown by the subject's ability to indicate subsequently that stimulus item to which each response had been given.

In addition to these main hypotheses, the design allowed a test as to whether responses made following the presentation of particular items would be matched against synonymous items, thus indicating that the original response had been a response to *meaning*.

For the experiment, seven subjects were each presented at a subliminal intensity with 12 items of stimulus material, dichotomized regarding their emotional values, in four consecutive trials. GSRs, response latencies, and verbal guesses were recorded for each presentation. Subsequently, each subject was presented with his responses and asked to match these against the 12 stimulus items.

1. Little evidence was found to support the hypothesis that response latency would be a function of the emotional value of the stimulus material.

2. The GSR hypothesis, however, was supported at a significant level of probability.

3. That the significant relationship between GSRs and the stimulus material justified the assumption of a causal relationship between the subliminal stimulus and the verbal response, received support from the second part of the experiment in which it was found that the group as a whole were able to match their verbal responses against the appropriate stimulus items to an extent that was statistically significant. That the responses were, in fact, responses to meaning was further supported by the subjects' ability to match them against the synonyms of the stimulus item by which they had been evoked. The results as a whole were considered to be supportive of the subception hypothesis.

REFERENCES

1. Allport, F. H. Theories of perception and the concept of structure. New York: John Wiley, 1955.
2. Baker, L. E. The influence of subliminal stimuli upon verbal behavior. *J. exp. Psychol.*, 1937, **20**, 84–99.
3. Bennett, M. Perceptual defense and threshold measurements. *J. pers.*, 1953, **22**, 565–571.
4. Bexton, W. H., Heron, W., & Scott, J. H. Effects of decreased variation in the sensory environment. *Canad. J. Psychol.*, 1954, **8**, 70–76.
5. Bricker, P. D., & Chapanis, A. Do incorrectly perceived tachistoscopic stimuli convey some information? *Psychol. Rev.*, 1953, **60**, 181–188.
6. Dixon, N. F. An investigation of auditory perception when the stimulus is subliminal. Unpublished thesis, Reading Univer., 1953.
7. Dixon, N. F. Symbolic associations following subliminal stimulation. *Int. J. Psycho-Anal.*, 1956, XXXVII, 159–170.

8. DUNLAP, K. The effect of imperceptible shadows on the judgement of distance. *Psychol. Rev.*, 1900, **7,** 435–453.
9. ERIKSEN, C. W. Subception: Fact or artifact. *Psychol. Rev.*, 1956, **63,** 74–80.
10. FISHER, C. Dreams and perception. *J. Amer. psychoanal. Ass.*, 1954, **11,** 389–445.
11. HOWES, D. A statistical theory of the phenomenon of subception. *Psychol. Rev.*, 1954, **61,** 98–110.
12. JACOBS, A. Formation of new associations to words selected on the basis of reaction-time-GSR combinations. *J. abnorm. soc. Psychol.*, 1955, **51,** 371–377.
13. LAZARUS, R. S. Subception: Fact or artifact? A reply to Eriksen. *Psychol. Rev.*, 1956, **63,** 343–347.
14. LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness, a study of subception. *Psychol. Rev.*, 1951, **58,** 113–122.
15. MILLER, J. G. Discrimination without awareness. *Amer. J. Psychol.*, 1939, **52,** 562–578.
16. POETZL, O. Experimentelle Erregte Traumbiläer in ihren Beziehungen zum Indirekten Sehen. *Ztschr. F. Neurol. Psychiat.*, 1917, **37,** 278–349.
17. STEVENS, W. L. Tests of significance for extra sensory perception data. *Psychol. Rev.*, 1939, **46,** 142–150.
18. WILLIAMS, A. C. Perception of subliminal visual stimuli. *J. Psychol.*, 1938, **6,** 187–199.

Received June 11, 1957.

ATTITUDES ESTABLISHED BY CLASSICAL CONDITIONING[1]

ARTHUR W. STAATS AND CAROLYN K. STAATS
Arizona State College at Tempe

OSGOOD and Tannenbaum have stated, "... The *meaning* of a concept is its location in a space defined by some number of factors or dimensions, and *attitude* toward a concept is its projection onto one of these dimensions defined as 'evaluative' " (9, p. 42). Thus, attitudes evoked by concepts are considered part of the total meaning of the concepts.

A number of psychologists, such as Cofer and Foley (1), Mowrer (5), and Osgood (6, 7), to mention a few, view meaning as a response—an implicit response with cue functions which may mediate other responses. A very similar analysis has been made of the concept of attitudes by Doob, who states, " '*An attitude is an implicit response ... which is considered socially significant in the individual's society*' " (2, p. 144). Doob further emphasizes the learned character of attitudes and states, "The learning process, therefore, is crucial to an understanding of the behavior of attitudes" (2, p. 138). If attitudes are to be considered responses, then the learning process should be the same as for other responses. As an example, the principles of classical conditioning should apply to attitudes.

The present authors (12), in three experiments, recently conditioned the evaluative, potency, and activity components of word meaning found by Osgood and Suci (8) to contiguously presented nonsense syllables. The results supported the conception that meaning is a response and, further, indicated that word meaning is composed of components which can be separately conditioned.

The present study extends the original experiments by studying the formation of attitudes (evaluative meaning) to socially significant verbal stimuli through classical conditioning. The socially significant verbal stimuli were national names and familiar masculine names. Both of these types of stimuli, unlike nonsense syllables, would be expected to evoke attitudinal responses on the basis of the pre-experimental experience of the *S*s. Thus, the purpose of the present study is to test the hypothesis that attitudes already elicited by socially significant verbal stimuli can be changed through classical conditioning, using other words as unconditioned stimuli.

METHOD

Subjects

Ninety-three students in elementary psychology participated in the experiments as *S*s to fulfill a course requirement.

Procedure

The general procedure employed was the same as in the previous study of the authors (12).

Experiment I.—The procedures were administered to the *S*s in groups. There were two groups with one half of the *S*s in each group. Two types of stimuli were used: national names which were presented by slide projection on a screen (*CS* words) and words which were presented orally by the *E* (*US* words), with *S*s required to repeat the word aloud immediately after *E* had pronounced it. Ostensibly, *S*s' task was to separately learn the verbal stimuli simultaneously presented in the two different ways.

Two tasks were first presented to train the *S*s in the procedure and to orient them properly for the phase of the experiment where the hypotheses were tested. The first task was to learn five visually presented national names, each shown four times, in random order. *S*s' learning was tested by recall. The second task was to learn 33 auditorily presented words. *S*s repeated each word aloud after *E*. *S*s were tested by presenting 12 pairs of words. One of each pair was a word that had just been presented, and *S*s were to recognize which one.

The *S*s were then told that the primary purpose of the experiment was to study "how both of these types of learning take place together—the effect that one has upon the other, and so on." Six new national names were used for visual presentation: *German*, *Swedish*, *Italian*, *French*, *Dutch*, and *Greek* served as the *CS*s. These names were presented in random order, with exposures of five sec. Approximately one sec. after the *CS* name appeared on the screen, *E* pronounced the *US* word with which it was paired. The intervals between exposures were less than one sec. *S*s were told they could learn the visually presented names by just looking at them but that they should simultaneously concentrate on pronouncing the auditorily presented words aloud and to themselves, since there would be many of these words, each presented only once.

---

[1] This study is part of a series of studies of verbal behavior being conducted by the authors at Arizona State College at Tempe. The project is sponsored by the Office of Naval Research (Contract Number NONR-2305 (00)), Arthur W. Staats, principal investigator.

The names were each visually presented 18 times in random order, though never more than twice in succession, so that no systematic associations were formed between them. On each presentation, the *CS* name was paired with a different auditorily presented word, i.e., there were 18 conditioning trials. *CS* names were never paired with *US* words more than once so that stable associations were not formed between them. Thus, 108 different *US* words were used. The *CS* names, *Swedish* and *Dutch*, were always paired with *US* words with evaluative meaning. The other four *CS* names were paired with words which had no systematic meaning, e.g., chair, with, twelve. For Group 1, *Dutch* was paired with different words which had positive evaluative meaning, e.g., gift, sacred, happy; and *Swedish* was paired with words which had negative evaluative meaning, e.g., bitter, ugly, failure.[2] For Group 2, the order of *Dutch* and *Swedish* was reversed so that *Dutch* was paired with words with negative evaluative meaning and *Swedish* with positive meaning words.

When the conditioning phase was completed, *S*s were told that *E* first wished to find out how many of the visually presented words they remembered. At the same time, they were told, it would be necessary to find out how they *felt* about the words since that might have affected how the words were learned. Each *S* was given a small booklet in which there were six pages. On each page was printed one of the six names and a semantic differential scale. The scale was the seven-point scale of Osgood and Suci (8), with the continuum from pleasant to unpleasant. An example is as follows:

German
pleasant:__:__:__:__:__:__:__:unpleasant

The *S*s were told how to mark the scale and to indicate at the bottom of the page whether or not the word was one that had been presented.

The *S*s were then tested on the auditorily presented words. Finally, they were asked to write down anything they had thought about the experiment, especially the purpose of it, and so on, or anything they had thought of during the experiment. It was explained that this might have affected the way they had learned.

Experiment II.—The procedure was exactly repeated with another group of *S*s except for the *CS* names. The names used were *Harry*, *Tom*, *Jim*, *Ralph*, *Bill*, and *Bob*. Again, half of the *S*s were in Group 1 and half in Group 2. For Group 1, *Tom* was paired with positive evaluative words and *Bill* with negative words. For Group 2 this was reversed. The semantic differential booklet was also the same except for the *CS* names.

Design

The data for the two experiments were treated in the same manner. Three variables were involved in the design: conditioned meaning (pleasant and unpleasant); *CS* names (*Dutch* and *Swedish*, or *Tom* and *Bill*); and groups (1 and 2). The scores on the semantic differential given to each of the two *CS* words were analyzed in a 2 x 2 latin square as described by Lindquist (4, p. 278) for his Type II design.

RESULTS

The 17 *S*s who indicated they were aware of either of the systematic name-word relationships were excluded from the analysis. This was done to prevent the interpretation that the conditioning of attitudes depended upon awareness. In order to maintain a counterbalanced design when these *S*s were excluded, four *S*s were randomly eliminated from the analysis. The resulting *N*s were as follows: 24 in Experiment I and 48 in Experiment II.

Table 1 presents the means and *SD*s of the meaning scores for Experiments I and II. The table itself is a representation of the 2 × 2 design for each experiment. The pleasant

TABLE 1

MEANS AND *SD*s OF CONDITIONED ATTITUDE SCORES

		Names			
		Dutch		*Swedish*	
Experiment	Group	Mean	*SD*	Mean	*SD*
I	1	2.67	.94	3.42	1.50
	2	2.67	1.31	1.83	.90
		Tom		*Bill*	
Experiment	Group	Mean	*SD*	Mean	*SD*
II	1	2.71	2.01	4.12	2.04
	2	3.42	2.55	1.79	1.07

Note.—On the scales, pleasant is 1, unpleasant 7.

TABLE 2

SUMMARY OF THE RESULTS OF THE ANALYSIS OF VARIANCE FOR EACH EXPERIMENT

Source	Exp. I			Exp. II		
	df	MS	*F*	*df*	MS	*F*
Between *S*s						
Groups	1	7.52	4.36*	1	15.84	5.00*
Error	22	1.73		46	3.17	
Within						
Conditioned attitude	1	7.52	5.52*	1	55.51	10.47**
Names	1	.02	.01	1	.26	.05
Residual	22	1.36		46	5.30	
Total	47			95		

* $p < .05$.
** $p < .01$.

---

[2] The complete list of *CS-US* word pairs is not presented here, but it has been deposited with the American Documentation Institute. Order Document No. 5463 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

extreme of the evaluative scale was scored 1, the unpleasant 7.

The analysis of the data for both experiments is presented in Table 2. The results of the analysis indicate that the conditioning occurred in both cases. In Experiment I, the *F* for the conditioned attitudes was significant at better than the .05 level. In Experiment II, the *F* for the conditioned attitudes was significant at better than the .01 level. In both experiments the *F* for the groups variable was significant at the .05 level.

DISCUSSION

It was possible to condition the attitude component of the total meaning responses of *US* words to socially significant verbal stimuli, without *S*s' awareness. This conception is schematized in Fig. 1, and in so doing, the way the conditioning in this study was thought to have taken place is shown more specifically. The national name *Dutch*, in this example, is presented prior to the word *pretty*. *Pretty* elicits a meaning response. This is schematized in the figure as two component responses; an evaluative response r_{PV} (in this example, the words have a positive value), and the other distinctive responses that characterize the meaning of the word, R_P. The pairing of *Dutch* and *pretty* results in associations between *Dutch* and r_{PV}, and *Dutch* and R_P. In the following presentations of *Dutch* and the words *sweet* and *healthy*, the association between *Dutch* and r_{PV} is further strengthened. This is not the case with associations R_P, R_S, and R_H, since they occur only once and are followed by other associations which are inhibitory. The direct associations indicated in the figure between the name and the individual words would also in this way be inhibited.

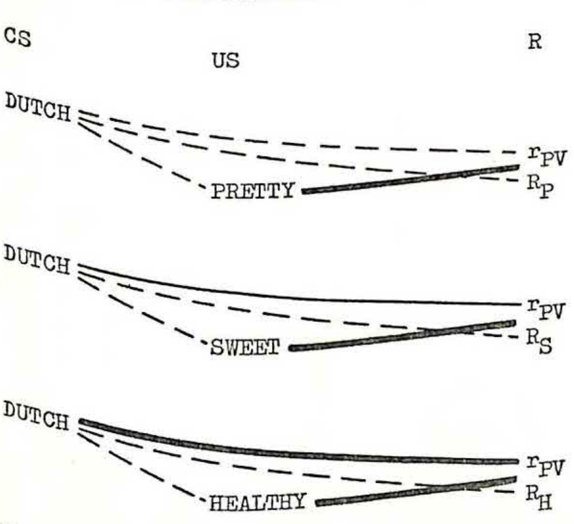

FIG. 1. THE CONDITIONING OF A POSITIVE ATTITUDE. THE HEAVINESS OF LINE REPRESENTS STRENGTH OF ASSOCIATION

It was not thought that a rating response was conditioned in this procedure but rather an implicit attitudinal response which mediated the behavior of scoring the semantic differential scale. It is possible, with this conception, to interpret two studies by Razran (10, 11) which concern the conditioning of ratings. Razran found that ratings of ethnically labeled pictures of girls and sociopolitical slogans could be changed by showing these stimuli while *S*s were consuming a free lunch and, in the case of the slogans, while the *S*s were presented with unpleasant olfactory stimulation. The change in ratings could be thought to be due to the conditioning of an implicit evaluative response, an attitude, to the *CS*s by means of the lunch or the unpleasant odors. That is, part of the total response elicited by the food, for example, was conditioned to the pictures or slogans and became the mediation process which in turn elicited the positive rating.

It should be stated that the results of the present study do not show directly that *S*s' behavior to the object (e.g., a person of Dutch nationality) has been changed. The results pertain to the *S*s' attitudinal response to the signs, the national names themselves. However, Kapustnik (3) has demonstrated that a response generalized to an object when the response had previously been conditioned to the verbal sign of the object. Osgood states,

> The aggressive reactions associated with *Nazi* and *Jap* on a verbal level certainly transferred to the social objects represented under appropriate conditions. Similarly, prejudicial behaviors established while reading about a member of a social class can transfer to the class as a whole . . . (7, p. 704).

The results of this study have special relevance for an understanding of attitude formation and change by means of verbal communication. Using a conception of meaning as a mediating response, Mowrer (5) has suggested that a sentence is a conditioning device and that communication takes place when the meaning response which has been elicited by the predicate is conditioned to the subject of the sentence. The results of the present study and the previous one of the present authors

(12) substantiate Mowrer's approach by substantiating the basic theory that word meaning will indeed condition to contiguously presented verbal stimuli. In the present study, the meaning component was evaluative, or attitudinal, and the *CS*s were socially significant verbal stimuli. The results suggest, therefore, that attitude formation or change through communication takes place according to these principles of conditioning. As an example, the sentence, "Dutch people are honest," would condition the positive attitude elicited by "honest" to "Dutch"—and presumably to any person called "Dutch." If, in an individual's history, many words eliciting a positive attitude were paired with "Dutch," then a very positive attitude toward this nationality would arise.

The reason for the group differences in each of the experiments is not clear. These differences could have arisen because there were actual differences in the *S*s composing each group, or in some condition of the procedure occurring to one of the groups. Nothing the authors were aware of seem to indicate this as the explanation, and in the previous experiments of the authors (12) there were no group differences. Since in a 2 x 2 latin square the interactions are entirely confounded with the main effects, the group differences could also have arisen as a result of the interaction of the other two main effects (i.e., direction of conditioning and names).

SUMMARY

Two experiments were conducted to test the hypothesis that attitude responses elicited by a word can be conditioned to a contiguously presented socially significant verbal stimulus. A name (e.g., *Dutch*) was presented 18 times, each time paired with the auditory presentation of a different word. While these words were different, they all had an identical evaluative meaning component. In Experiment I, one national name was paired with positive evaluative meaning and another was paired with negative evaluative meaning. In Experiment II, familiar masculine names were used. In each experiment there was significant evidence that meaning responses had been conditioned to the names without *S*s' awareness.

REFERENCES

1. COFER, C. N., & FOLEY, J. P. Mediated generalization and the interpretation of verbal behavior: I. Prologemena. *Psychol. Rev.*, 1942, **49**, 513-540.
2. DOOB, L. W. The behavior of attitudes. *Psychol. Rev.*, 1947, **54**, 135–156.
3. KAPUSTNIK, O. P. The interrelation between direct conditioned stimuli and their verbal symbols. (Trans. from Russian title) *Psychol. Abstr.*, 1934, **8**, No. 153.
4. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
5. MOWRER, O. H. The psychologist looks at language. *Amer. Psychologist*, 1954, **9**, 660–694.
6. OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 197–237.
7. OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford Univer. Press, 1953.
8. OSGOOD, C. E., & SUCI, G. J. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50**, 325–338.
9. OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychol. Rev.*, 1955, **62**, 42–55.
10. RAZRAN, G. H. S. Conditioning away social bias by the luncheon technique. *Psychol. Bull.*, 1938, **35**, 693.
11. RAZRAN, G. H. S. Conditioned response changes in rating and appraising sociopolitical slogans. *Psychol. Bull.*, 1940, **37**, 481.
12. STAATS, C. K., & STAATS, A. W. Meaning established by classical conditioning. *J. exp. Psychol.*, 1957, **54**, 74–80.

Received June 12, 1957.

CHANGES IN RELIGIOUS INTEREST: A RETEST AFTER 15 YEARS

IRVING E. BENDER
Dartmouth College

ABUNDANT as are follow-up studies of college students, all too few have been concerned with a program of research during, as well as after, college; particularly is this the case in respect to research on motives and values. This report represents an attempt to capitalize on the opportunity for such research presented by the fact that in 1939–40 an intensive psychological study (4) has been made of 124 seniors at Dartmouth College by means of tests, ratings, interviews, and autobiographies. Although these men, one quarter of the senior class, were selected for a variety of reasons, the sample was found to be representative of the whole class in terms of scholastic aptitude and performance (4, p. 48).

A renewed study of these same men was made possible in 1955–56 through a grant from the American Philosophical Society; 61 men were interviewed and 84 were retested on the Study of Values. Also, a questionnaire was mailed to the 118 surviving subjects of the original study, and, by tugging and pulling, a final count of 112 forms were returned—a 95 per cent triumph.

A few general characteristics of this group may be mentioned. In 1956, they ranged in age from 36 to 42. Almost all the men are married, only four have never married. Five men have been divorced, four of whom have remarried. Of the married group, all but six have children, with an average of 2.5 per family. However, twelve families have four children and two families have five. Weekly church attendance is reported by 38 men, but 24 do not go at all. The remainder go sometimes. Well over half the men report that their children attend Sunday school regularly. The group is predominantly Protestant. Eleven are Catholic and nine are Jewish. More than half of the men continued with graduate work after college and 47 have advanced degrees, chiefly in business and administration, but four have the Ph.D. degree. Of the 112 men, 76 are in business activities, 12 are teachers, 9 lawyers, 6 physicians, 4 writers, 1 chemist, 1 architect, 1 engineer, 1 minister, and 1 retired. In one of the earliest studies of Dartmouth College, King (7) reported that the 64 men in the class of 1880 were in the following vocations: 17 lawyers, 14 teachers, 11 businessmen, 7 physicians, 7 ministers, 4 engineers, 1 journalist, and 1 chemist (two men died soon after graduation). The present shift in vocational interest is apparent.

One of the measures that proved particularly helpful in the original study of motivation was the Allport-Vernon Study of Values, 1931 edition. Respecting subjective values, Allport writes, they are "the core of the dynamics of behavior, and play so large a part in unifying the personality" (2, p. 427). So a plan was arranged to use the same edition of the Study of Values, and 84 men were retested during the years of 1955 and 1956. It is to be recalled that these scores are relative; a high score in one or more values is counterbalanced by a low score in one or more of the remaining values. There are 180 points to be divided among six values; the theoretical mean of each is 30.

RESULTS AND DISCUSSION

Three main findings of this study are reported. First, a general survey of the shifts in the six values from 1940 to 1956; secondly, the correlations of religious scores and other variables; and thirdly, an analysis of the religious-scoring items of the test.

Shifts in Value Scores

Table 1 reports the mean scores and the t values of the differences between the 1939 and 1940 scores on the Study of Values and the 1955 and 1956 scores; the former scores are indicated as T40 and the latter as T56 scores.

The political value—interest in power and influence—remains the highest value of these respondents, only slightly lower after the lapse of 15 years. The most striking change is the highly significant increase in the religious value, which has now risen from its subbasement position in 1940 to become a close runner-up to the political value in 1955–56. Higher religious scores in the retest occur for 80 per cent of the group. Of the 118 men, only

TABLE 1

COMPARISON OF ALLPORT-VERNON SCORES, 1939–40 vs. 1955–56
($N = 84$)

Value	T40 Scores		T56 Scores			t
	Mean	SD	Mean	SD	$r_{T40-T56}$	
Theoretical	28.30	7.09	30.29	5.84	+.48	2.69**
Economic	31.51	9.03	29.63	9.18	+.53	1.92
Aesthetic	30.67	8.14	25.81	9.48	+.61	5.58**[a]
Social	30.88	6.15	28.98	5.52	+.20	2.34*
Political	33.44	6.90	32.80	6.89	+.41	.77
Religious	25.21	8.28	32.37	8.93	+.49	7.43**[a]

* $p = .05$ at $t = 1.99$.

** $p = .01$ at $t = 2.64$.

[a] Kelly (6) reports the following data for value change among 176 men over 20 years:

Value	Time 1		Time 2			
	M	SD	M	SD	r	t
Aesthetic	28.6	7.9	26.0	7.4	+.32	4.7
Religious	29.0	9.1	33.8	9.1	+.60	7.5

one is a minister. As expected, his highest scores are in the religious value (T40: 40; T56: 39). However, of the 84 men who were retested, 22 men or 26 per cent score at least as high as the minister, while in 1940, only 4 per cent scored as high as the minister-to-be's T40 score. The religious value is the highest score for 23 of the retested group, 27 per cent of the respondents. In 1940, the religious value was highest for only 6 of these 84 men, 7 per cent.

In all values except the social, T40 and T56 scores are significantly correlated (Table 1). The social value has been frequently shown to fall short of reliability. The conclusions of the present study rest of course on the degree of reliability and validity that has been demonstrated for the Study of Values (5, 8, 9).

In Table 1, differences between T40 and T56 mean scores of the matched groups are evaluated by t test. Religious, aesthetic, and theoretical values have all changed significantly. Since the most arresting change is in the religious value, the question arises if this increment is reliably greater than the other changes; particularly, is it greater than for the marked drop in the aesthetic value? Since the difference between the change in the mean religious scores and that in mean aesthetic scores has a t value of 4.8 ($p < .01$), the shift to religious interest is decisively predominant. Kelly (6) reports a similar finding for 176 men after 20 years; the most significant shifts in the Allport-Vernon test were the increase in the religious value and a decrease in the aesthetic value. (See footnote to Table 1.)

The aesthetic value, "the search for form and harmony," has never been salient for these men; at best in 1940 it occupied only a middle position. Perhaps during college more than now, slightly greater opportunities presented themselves to enjoy "each single impression . . . for its own sake" (2), but at present this attitude is at low ebb.

The Study of Values is based on Spranger's *Lebensformen*, where the religious man is said "to endow the world with meaning and value . . . he finds something divine in every aspect of life" (1, p. 212–213). The religious value implies an immanent faith in a higher reality than that of everyday life, a reality concerned with the experience of unity. As we shall see, questions in the test that relate to this attitude have changed less than those which have to do with more orthodox views about the divine being.

Correlates of the Religious Value

Our second consideration deals with correlations of variables with religious value scores in 1940 and 1956, reported in Table 2. The r of +.79 between 1956 scores and church attendance supports the validity of the religious score.

During the investigation of motivation of these men in 1939–40, the men were rated on the following scale of Value-Energy:

1A Inspired: Is fired with enduring zeal and is not to be discouraged or deflected from his course.
1B Intensive: Has a firm faith in the importance of his values and is motivated to follow them out steadily and eagerly.
2 Seeking: Is trying earnestly to distill values from his experiences which will serve to centralize his purposes.
3 Anxious: Is motivated to compete by his desire to display and sometimes tries to do more than his potentialities warrant.
4 Vague: Has no strong interests or values of his own but depends upon external motivation such as rigid requirements or fear of punishment.
4A Drifting: Aimless in values and tending to derive his motivation from momentary stimulation.
5 Abulic: Is unable to make up his mind about any values and almost totally lacking in motivation.

In 1955–56, the men who were interviewed were rated again in Value-Energy. An r of

TABLE 2
CORRELATIONS OF RELIGIOUS SCORES IN THE STUDY OF VALUES AND OTHER VARIABLES

Variable	T40 Scores		T56 Scores	
	r	*N*	*r*	*N*
Church attendance			+.79**	82
Children's Church attendance			+.24	62
Value-Energy rating 1955–56	+.14	57	+.36*	51
Value-Energy rating 1940	−.12	107	−.13	51
Amer. psychol. exam.	−.14	106	−.16	82
No. of extra-curricular activities	−.05	100	−.02	80
Extroversion Scale A-B	−.38**	54	−.29*	45
Means of professorial ratings	−.02	102	+.04	80
Self-Development ratings 1956	+.12	49	+.30*	44

* $p < .05$.
** $p < .01$.

+.31 is found between the ratings of Value-Energy of 1940 and 1955–56. This correlation is to be interpreted as not so much a measure of reliability as a reflection of some degree of consistency on the part of those men in following out their purposes from college to their present situations. If these ratings are valid, Value-Energy tends to have an enduring quality. The 1940 ratings of Value-Energy are not significantly related to the religious scores. But the new 1956 ratings correlate at the .05 level with the 1956 religious scores. Nine men are rated lower in Value-Energy in 1956 than in 1940. The remaining 52 men compared with this small group have reliably higher scores in the religious value.

In 1955 and 1956, after interviewing the men, they were also rated in the variable called Self-Development, the degree to which they were living out their purposes and potentialities. The scale for rating Self-Development grew out of a series of interview questions aimed at determining the respondent's (*a*) goal attainment, whether or not he was getting from life what *he* wanted and also (*b*) self-actualization, to what extent he was using his capacities and inner resources as determined by the author's knowledge of tests taken during his college career. Separate ratings on a five-step scale were made of each of these variables. Since an *r* of +.72 was found between the two sets of ratings, they were combined and designated as ratings of Self-Development.

These Self-Development ratings correlate at the .05 level with the 1956 religious scores. Since the author also rated the men in terms of how well he liked them, a favorable attitude may well have impinged upon the ratings of Value-Energy and those of Self-Development. However, the chance relationship between liking and religious scores tends to deny such an implication. The only other significant correlations concern the negative relationship between the 1940 ratings of extroversion and religious scores T40 and T56.

Further light on the changes in religious values is cast by an examination of scores by different subgroups of respondents. In Table 3, the results are reported for the men in the three divisions of general study, namely, the humanities, sciences, and social sciences. The increase of 9.2 points from 1940 to 1956 in the religious score of the 21 men who had been students of the humanities approaches significance at the .05 level. The increase of 6 points from 1940 to 1956 scores for the 52 social sciences men is, however, highly significant. When the mean differences between

TABLE 3
COMPARISON OF MEAN RELIGIOUS VALUE SCORES (T40 VS. T56) FOR MAJOR DIVISIONS OF STUDY AT COLLEGE

Major Divisons	*N*	T40 Scores		T56 Scores		*t*
		Mean	*SD*	Mean	*SD*	
1 Humanities	21	25.5	7.5	34.7	10.0	1.92*
2 Science	11	25.4	9.3	30.2	9.7	.83
3 Social science	52	24.5	8.0	30.5	5.7	4.6**

* $p < .05$.
** $p < .01$.

TABLE 4
COMPARISON OF RELIGIOUS VALUE SCORES (T56) OF PRESENT OCCUPATIONAL GROUPS

	N	Mean	*SD*
1 Executives in large corporations	20	29.65	9.0
2 Small self-run business	10	38.5	9.0
3 Family business	14	31.3	6.1
4 Salesmen	11	33.0	7.5
5 Salaried professionals	19	31.5	8.7
6 Independent professionals	9	33.4	11.3

t Values Between the Groups

	1	2	3	4	5	6
1		2.5*	.61	1.1	.66	.83
2			2.66*	1.45	1.66	1.02
3				.56	.07	.46
4					.49	.14
5						.79

* $p = .02$.

TABLE 5

COMPARISON OF MEAN RELIGIOUS VALUE SCORES (T40 VS. T56) FOR RELIGIOUS DENOMINATIONS

Denomination	T40 Scores			T56 Scores			t
	Mean	N	SD	Mean	N	SD	
Congregational	24.0	25	7.7	29.4	20	7.7	2.35*
Episcopal	23.4	22	7.9	30.3	17	7.9	2.56*
Presbyterian	24.6	16	8.2	34.0	12	9.8	3.61**

* $p < .05$.
** $p < .01$.

TABLE 6

RELIGIOUS ITEMS SHOWING SIGNIFICANT CHANGE FROM T40 TO T56

Item No. in The Study of Values	Mean Item Score		t
	T40	T56	
Part I			
6. Which of these character traits do you consider the most desirable: (*a*) *high ideals and reverence;* (b) unselfishness and sympathy?	.8	1.3	3.60
13. All the evidence that has been impartially accumulated goes to show that the universe has evolved to its present state in accordance with mechanistic principles, so that there is no need to assume a first, cosmic purpose, or God behind it, (a) Yes; (*b*) *No.*	1.43	2.47	6.0
21. The aim of the churches at the present time should be: (a) to bring out altruistic and charitable tendencies, and to urge people to think more of the good of others; (*b*) *to convey spiritual worship, and a sense of communion with the highest.*	.60	1.08	3.42
26. Taking the Bible as a whole, one should regard it from the point of view of its beautiful mythology and literary style rather than as a spiritual revelation. (a) Yes; (*b*) *No.*	1.04	1.86	5.60
Part II			
2. In your opinion, can a man who works in business for his living all the week best spend Sunday in: a. trying to educate himself by reading serious books b. trying to win at golf or racing c. going to an orchestral concert *d. hearing a really good sermon*	3.13	2.0	6.72
5. If you lived in a small town and had more than enough income for your needs, would you prefer to: a. apply it productively to industrial development *b. help to endow the church to which you belong* c. give it to a university for the development of scientific research d. devote it to hospitals	3.52	2.72	5.3
7. Assuming that you are a man with the necessary ability, and that the salary for each of the following occupations is the same, would you prefer to be a: a. mathematician b. sales manager *c. clergyman* d. politician	3.2	2.82	3.16
12. Should one guide one's conduct according to, or develop one's chief loyalties toward: *a. one's religious faith* b. ideals of beauty c. one's business organizations and associates d. society as a whole	3.14	2.33	5.06
14. If you should marry (or are married) do you prefer a wife who: a. can achieve social prestige, commanding admiration from others b. likes to stay at home and keep house *c. is fundamentally spiritual in her attitude toward life* d. is gifted along artistic lines	3.20	2.57	3.93

the diversions are studied, none is significant, either for 1940 or 1956.

Robert Gutman of the Sociology Department of Dartmouth College set up occupational groupings of the men for whom 1956 value scores were available. The mean differences are shown in Table 4. The outstanding differences occur for the self-initiated small business group who show significantly higher T56 scores in religious values than do the large corporation executives or the family business group.

Table 5 gives findings according to the former student's religious denomination as of the time he entered college. Differences in mean T40 and T56 scores are reported for the three denominations represented by more than ten cases, and the differences are found to be reliable. No interdenominational differences of the means for either 1940 or 1956 are reliably different.

A few additional facts deserve mention. A comparison of 38 men who report weekly church attendance and the 24 who do not attend at all indicates that the latter group report reliably lower goal achievement and less civic participation, although religious value as such did not turn out to be reliably correlated with the latter variables. Only 17 men of the 84 retested decreased their religious score. This deviant group is reliably different ($p < .05$) from the larger group in their 1956 religious scores. They also report lower goal achievement, less participation in civic affairs, and less frequent church attendance. Of the six men from whom it proved impossible to obtain returns, the mean religious score was 20 in 1940 compared with the total 1940 mean of 25.

Item Changes

We have seen a marked increase in mean religious value score for Dartmouth men over 15 years. In what specific ways did these men change their religious values? Was the change based on a more mature philosophy of life or was it in a more theological direction?

For all but three of the 20 religious terms in the Allport-Vernon scale, change is in the direction of religious interest; significantly so at the .01 level in the case of nine of the items (Table 6). A study of the nature of the significant changes points up the quality of the upsurge in the religious value over the period of 15 years. The turn in the religious direction seems more theological than philosophical.

The format of the Allport-Vernon instrument results in the fact that each value is compared twice with each of the other values. A study of such comparisons shows that there are reliable differences in favor of religious interest in both instances when the religious value is matched with the aesthetic and with the social values and in one instance only when matched with theoretical, economic, and political values.

Kelly (6) has raised the question of whether the increase in religious value was a function of the person's maturity or a function of the times in which we live. To study this problem, the same test form was given to 66 present undergraduates, mostly seniors. On comparison of present undergraduates to the seniors of 1940 (Table 7), a reliably higher religious mean score is found for present undergraduates and a reliably lower mean economic score. Comparison of the mean scores of present undergraduates with the mean scores attained by the students of 1940 as retested in 1956, however, reveals a remarkable similarity of values. These data would suggest that the temper of the times in which we live influences the religious value more than does the maturity of the men. Apparently, the same need for religious interest exists now among the young as among the older.

TABLE 7
COMPARISON OF PRESENT AND FORMER DARTMOUTH STUDENTS WITH RESPECT TO ALLPORT-VERNON SCORES

Value	Former Students		Present Students		Comparisons	
	(*N* = 84)		(*N* = 66)		*t*	*t*
	Mean[a] T40 (1)	Mean[a] T56 (2)	Mean	*SD*	$\bar{1}$ vs. 3	$\bar{2}$ vs. 3
Theoretical	28.30	30.29	30.41	7.45	1.74	.01
Economic	31.51	29.63	28.21	8.91	2.23*	.94
Aesthetic	30.67	25.81	28.33	8.69	1.69	1.68
Social	30.88	28.98	29.17	5.40	1.80	.02
Political	33.44	32.80	32.21	5.81	1.17	.56
Religious	25.21	32.37	31.89	9.46	4.48**	.31

** = .01 at 2.60.
* = .05 at 1.97.
[a] See Table 1 for SDs.

From interviews with the men, it becomes clear that religion in one form or another has usually become important to them. Furthermore, in a number of cases it has given fuller meaning and purpose to life. The interviews confirm Allport's statement: "While religion certainly fortifies the individual against the inroads of anxiety, doubt, and despair, it also provides the forward intention that enables him at each stage of his becoming to relate himself meaningfully to the totality of being" (3, p. 96).

Interesting, indeed, are the *speculative* questions that this change in religious interest suggests. Is the resurgence of religion a part of the widespread and expanding need to belong, the need to conform? Or is the change motivated by fear, not without an undertone of hope? Do circumstances compel one to despair of reason and invoke the ancient answer to trust in God?

From the interviews with more than half of the men either in their homes or in their offices, it was apparent that they are by and large successful socially, vocationally, and financially. Their families are loved and well provided for. However, what was patently sought was a perspective for something purposive beyond the material comforts of their existence. Certain vague anxieties were expressed about the routine and the literalness of their lives. Increasingly, the return to religion seems to quell their fears and to promise hope.

SUMMARY

College students who were originally studied in 1940 were interviewed and retested in 1955 and 1956. The results of changes in the religious value scores on the Allport-Vernon Study of Values are reported here. A significant increase in the religious value scores is found after the 15-year interval.

These religious scores correlate highly with church attendance and to a lesser extent with present ratings of Value-Energy and Self-Development. Dividing the respondents into subgroups, we find the most marked increase in religious value scores on the part of (*a*) respondents who were students in the social sciences, (*b*) those who are now operating a small self-run business, and (*c*) those who class themselves as Presbyterians. Among the religious scoring items, the theological items have increased more than the philosophical ones. The religious value scores of present undergraduates are reliably higher than for the men who took the test in 1940, and are remarkably similar to the respondents who took the test in 1956.

REFERENCES

1. ALLPORT, G. W. *Personality: A psychological interpretation.* New York: Holt, 1937.
2. ALLPORT, G. W., & VERNON, P. E. *A study of values: Manual of direction.* New York: Houghton Mifflin, 1931.
3. ALLPORT, G. W. *Becoming.* New Haven: Yale Univer. Press, 1955.
4. BENDER, I. E., *et al. Motivation and visual factors.* Hanover: Dartmouth Coll. Publ., 1942.
5. CANTRIL, H., & ALLPORT, G. W. Recent applications of the study of values. *J. abnorm. soc. Psychol.*, 1933–34, **28,** 259–279.
6. KELLEY, E. L. Consistency of the adult personality. *Amer. Psychologist*, 1955, **11,** 659–681.
7. KING, S. T. *Dartmouth College, 1880—Chronicles fifty years later—1930.* Milford, N. H.: Cabinet Press, 1930.
8. VERNON, P. E., & ALLPORT, G. W. A test for personal values. *J. abnorm. soc. Psychol.*, 1931–32, **26,** 231–248.
9. WHITELY, P. L. A study of the Allport-Vernon test for personal values. *J. abnorm. soc. Psychol.*, 1933–34, **28,** 6–13.
10. WHITELY, P. L. The constancy of personal values. *J. abnorm. soc. Psychol.*, 1938, **33,** 405–408.

Received June 14, 1957.

ATTEMPTS TO LEAD, GROUP PRODUCTIVITY, AND MORALE UNDER CONDITIONS OF ACCEPTANCE AND REJECTION[1]

PAULINE N. PEPINSKY,
Ohio State University
JOHN K. HEMPHILL,
Education Testing Service
AND REUBEN N. SHEVITZ
Eli Lilly Company

In everyday life it sometimes happens that the members of a group are faced with a contingency such that for any one of them to try to lead is to invite personal rejection by the rest. This situation arises when, as the members see it, what must intervene between them and task success is bound to be dangerous, obnoxious, exhausting, or generally unpleasant. Whoever proposes any task-relevant course of action is sure to be unpopular, and yet the outcome of the project depends upon the initiation of group action. Under these restrictive circumstances, who can afford to attempt to lead his group? What is the individual like who is willing and able to take this social risk? These are the central questions that provoked the experiment to be described.

In addition, supplementary data collected during the study provide us with information about productivity and morale under these special conditions. Ever since the pioneer experiment of Lewin, Lippitt, and White (11, 12) social "meteorologists" have been interested in the effects of leader behavior upon group "climate" and, also, in the effects of such atmospheric conditions upon group productivity. Our collateral analysis tells us something about what happens to group productivity and morale when the tables are turned, and group climate varies as a function of the behavior of *other* members *toward* those who try to assume leadership.

This experiment was designed primarily to test the hypothesis that when attempts to lead are repeatedly met by personal rejection, group members characterized by a relatively high need for achievement and a low need for affiliation will make such attempts more frequently than other members who have low needs for achievement, but relatively high affiliation needs. This prediction was based upon a provisional theory of leadership formulated by Hemphill (6), and the study is one of a series of four small-group experiments conducted to test some of the propositions stated in that theory. These studies (8, 9, 14, 18) all were concerned with the identification of motivational variables antecedent to *attempts* to lead, without regard for whether such attempts actually were followed by the group or did in fact lead to task success. In effect, the general question was, "What makes Johnny run," rather than, "What makes Johnny get elected." Throughout this exploratory research the major dependent variable was the observed frequency of "attempted leadership acts," defined by Hemphill (6) as attempts to "initiate structure-in-interaction" through making task-relevant proposals or suggestions (cf. Bass *et al.*, 3).

In this particular experiment, the independent variables were (*a*) the inferred needs of the subjects, as measured prior to the experiment, and (*b*) the special conditions to which they were then exposed. The conceptions of needs for achievement and affiliation used here are similar to but not identical with those proposed by McClelland *et al.* (13), and by Shipley and Veroff (19). In Hemphill's view—

> *Need Achievement* (n Ach) is inferred when observed behavior is interpreted as a consistent attempt by the individual to experience success in competition with some standard of excellence.

---

[1] This research was conducted under Contract N6ori-17 T.O. III (NR 171-123) between the Ohio State University Research Foundation and the U. S. Office of Naval Research. The opinions expressed are those of the authors alone. They wish to thank the following persons for their special assistance: Charlotte A. Christner, A. E. Kaufman, and M. E. Lipetz, graduate assistants, who participated in various phases of the research; H. J. Ehrlich, who served as an observer; J. R. Hanson, who functioned as one of the trained "stooges"; and H. B. Pepinsky, who gave constructive criticism in the preparation of the manuscript. Sample copies of the forms used in the study are appended to the original technical report to ONR (14), which may be obtained on loan from the Gifts and Exchange Department of The Ohio State University Library.

Need Affiliation (n Aff) is inferred when observed behavior is interpreted as a consistent attempt by the individual to establish, maintain, or recover friendly, warm, or loving relationships with other persons.

Hemphill introduces an emphasis upon consistency in behavior in different situations and over a period of time and regards such consistency as an index of the strength of the respective needs. In contrast, McClelland's theoretical position has led to the development of measurement procedures that are critically dependent upon arousal in a specific situation.

The situational arrangements in the present study were designed to establish one or the other of two conditions, but otherwise to maintain equal opportunities for task success:

The rejection condition (*R*) was intended to produce in the *S*s the expectation that attempting to lead would result in their personal rejection by other group members. (Under this condition, attempting to lead was assumed to have satisfying consequences for need Achievement, but dissatisfying consequences for need Affiliation.)

The acceptance condition (*A*) was intended to produce in the *S*s the expectation that attempting to lead would result in their personal acceptance by other group members. (Under this condition, attempting to lead was assumed to have satisfying consequences for both need Achievement and need Affiliation.)

Strictly interpreted, the stated hypothesis refers to behavior to be observed under condition R only. But by adding condition A to the design, we could extend that information by relating it to what happened in a situation in which differences in needs were *not* expected to have a significant effect upon attempts to lead.

METHOD

Selection of Subjects

By means of multiple screening procedures, 48 white male *S*s were selected from the students enrolled in the University introductory psychology course and assigned to 24 four-man groups, each comprised of two *S*s and two "stooges." The major criterion for selection of the *S*s was their classification in respect to n Ach and n Aff. This classification was accomplished in two stages: first, potential *S*s were prescreened according to their scores on a questionnaire, and, second, final selection was made in terms of independent ratings of their needs, based upon individual, structured interviews.

The screening questionnaire consisted of two parts, a section that elicited general information and a series of items designed to measure the two needs. The first set of items permitted elimination of those individuals who did not meet criteria established to control the extraneous effects of status-related physical characteristics (e.g., age, physical size, and race). Items contained in the second section of the questionnaire were developed from an original pool of more than 400 statements, which the individual members of the research staff devised as descriptive of behavior consistent with *either* high or low n Ach or n Aff. A preliminary 30-item form (15 items intended to measure each need) then was assembled from those items that at least five of six independent judges agreed were clearly relevant to one or the other of the given need definitions.

On the basis of a trial administration to 161 men enrolled in the introductory psychology course, this section of the questionnaire was revised in order to eliminate nondiscriminating or apparently ambiguous items. In its final form the questionnaire contained 15 items, seven designed to yield a score for n Ach, and eight designed to yield a score for n Aff. The two scores are empirically independent; a correlation computed between them for a sample of 335 men is −.02. From a scatter plot of the distributions of the two scores, 86 individuals whose scores fell in the high-low quadrant—and who also met the physical criteria—were selected for interviewing.

Individual interviews, which averaged about 15 minutes in length, were conducted with the potential *S*s by one or the other of two trained interviewers. The interviewer had no knowledge of the individual's scores on the screening questionnaire, beyond the fact that his scores were extreme enough to qualify him for the interview. The specific questions asked (modifying a schedule used by Rafferty [15]) focused upon activities and interests (vocational, academic, social) that presumably would provide some basis for judgments as to the consistency with which, in various situations and over an extended time period, an individual had engaged in behavior symptomatic of n Ach or n Aff. Immediately after his interview, the interviewer rated each person on the two needs on separate nine-point scales. For any individual to be eligible to serve as an *S*, it was stipulated (*a*) that there should be a discrepancy of at least two scale intervals between his ratings on the two needs, and (*b*) that the lower of the two ratings should not exceed the fifth or midpoint on the scale. If, in the interviewer's judgment, an individual had met all the specified selection criteria, he was assigned to a particular experimental group, each containing one *S* who was classified as relatively high on n Ach and low on n Aff and one classed as high n Aff, low n Ach. Final selection and classification were based solely on the interview.

Task

All groups worked on the same experimental task, which was designed to maximize the necessity for planning and coordinating group action, and, hence, to elicit a relatively large number of attempts to lead. The basic idea elaborated upon in the development of the task was suggested by the work of Guetzkow and his associates (5) with hypothetical business corporations. This task, the Manufacturing Problem, required the group members to organize as a toy manufacturing concern and to operate their business for maximum profit.

The physical arrangements for the task involved the use of four tables placed around the laboratory: (*a*) the "Supplier's" table, where a large quantity and variety of Tinkertoy parts were available in appropriately labeled boxes; (*b*) a display table upon which five complete Tinkertoy models were arranged (a "top," a "man," an "airplane," a "wagon," and a "ladder"); (*c*) the "Shop," a large table upon which the "products" were to be assembled, and upon which were arranged pads, pencils and order forms; and (*d*) a "Buyer's" table at which finished toys were to be sold.

At the outset, each group was given three dollars in cash with which to "set itself up in business" and was told that the members would be permitted to keep equal shares of whatever profit they made during the experiment. Itemized lists of supply costs and selling prices were distributed, and it was pointed out that both costs and prices would fluctuate every five minutes throughout the two 20-minute work periods. In order to buy parts, the group had to fill out and submit to the supplier itemized order forms, each signed by all four team members and accompanied by sufficient money to cover the particular order. Each work session was preceded by a planning period (the first, 10, and the second, five minutes long); the first work session and the second planning period were separated by a 10-minute break. After the standard instructions had been read by the experimenter, the organization and operation of the business were left completely in the hands of the group. During the special planning periods no actual production was permitted, but discussion and revision of plans could be continued during the work sessions.

Experimental Design and Procedure

During their work on the experimental task half of the 24 groups were subjected to condition R and the other half to condition A, the conditions being alternated in RAAR order. In order to establish the conditions, the same pair of accomplices was introduced into each experimental group to play preassigned roles. Use was also made of the immediate feedback to the two experimental *S*s in each group of the prearranged "results" of a sociometric questionnaire administered twice during the experiment. Two trained observers, who observed the group through a one-way mirror, tallied the frequency of the leadership acts attempted during the experimental sessions by each *S*. The obsevers were not informed about the need classification of the *S*s, and the observational procedure was specified by rules developed in conformance with the theoretical definition of attempted leadership. The following procedural sequence comprised the complete experimental period: (*a*) a 15-minute preliminary or "warm-up" session, (*b*) the first administration of the sociometric questionnaire, (*c*) the first session of the experiment proper, (*d*) the second administration of the sociometric questionnaire, and (*e*) the second and final experimental session.

The preliminary session, during which the group worked on a simple construction task, was intended primarily to provide a plausible basis for the first administration of the sociometric questionnaire. In this period, the stooges played relatively passive roles, but voluntarily undertook enough of the work of the group to prevent either one of the *S*s from establishing himself as the dominant leader. The sociometric questionnaire consisted of seven items of the "guess-who" type, which elicited choices based upon criteria both of liking for them as individuals and of preference for them as leaders. After a brief interval during which the results supposedly were actually tallied, identical reports were given privately to each *S*, and he was cautioned not to discuss his ratings with the other group members. The intensity of reported rejection or acceptance increased from the first to the second report, and acceptance and rejection were varied to the same degree under both conditions. The feedback was concluded with a suggestion by the experimenter of the appropriate inference to be drawn by the *S*, i.e., that he was being accepted (or rejected) *because* of the suggestions he was making to the group.

Under Condition R, both stooges were instructed to show interest in and willingness to work on the task as such, but to indicate personal rejection of either *S* whenever he made an attempt to lead. This role was reversed under Condition A; whenever an *S* attempted to lead, the stooges indicated personal acceptance and approval of the initiator himself. In both instances, the stooges were to comply with all direct requests or orders and were themselves to make no attempts to lead.

As soon as the experiment was over, the four group members (including the stooges, who simulated compliance) filled out a brief Postsession Questionnaire. The *S*s also completed the Group Dimensions Description Questionnaire (GDDQ) (Hemphill, 7) which yields scores on 13 dimensions, providing a "profile of an individual's orientation (perception and attitudes) toward a group." Finally, the research supervisor explained the experimental procedures to the *S*s and reassured them about the adequacy of their performance.

RESULTS

Checks on Procedures

Independent checks upon the experimental procedure provide information about (*a*) the reliability of the interview as a selection criterion, (*b*) the manipulation of the experimental conditions, and (*c*) the reliability of the measure of the dependent variable.

To obtain an estimate of interviewer agreement, one interviewer made independent need ratings based upon tape recordings of 32 interviews conducted by the other interviewer. Since the two raters were responding to comparable but not equivalent stimuli, the relationship between the two sets of ratings should be regarded as an attenuated index of interviewer agreement. Correlations of .67 between their ratings of n Aff and .55 between their ratings of n Ach indicate a significant and positive, but not high degree of association. These correlations are, however, comparable to those obtained in studies of agreement in clinical judgments (e.g., Hunt, Arnhoff, and

TABLE 1

ANALYSIS OF VARIANCE OF ATTEMPTED LEADERSHIP SCORES

Source of Variation	*df*	Variance	F_1	F_2
Between Conditions	1	130.67	76.42**	21.92**
Between Needs	1	0.38	—	—
Between Sessions	1	5.04	2.95	—
Between Blocks	11	4.86	2.84*	—
Conditions × Needs	1	1.49	—	—
Conditions × Sessions	1	0.67	—	—
Conditions × Blocks	11	5.35	3.13*	—
Needs × Sessions	1	0.37	—	—
Needs × Blocks	11	2.60	1.52	—
Sessions × Blocks	11	0.68	—	—
Conditions × Needs × Sessions	1	4.17	2.44	—
Conditions × Needs × Blocks	11	5.96	3.49*	—
Conditions × Blocks × Sessions	11	1.17	—	—
Needs × Sessions × Blocks	11	0.60	—	—
Residual	11	1.71	—	—
Total	95			

* Significant at .05 level.

** Significant at .01 level.

Cotton [10] and Ash [2], (Only two of these 32 individuals actually would have been placed in opposite categories by the two interviewers; as it happened, one of these *S*s was run under A and the other under R.)

Two months after the completion of the experiment and more than four months after the initial interview ratings had been made, a third interviewer held similarly structured interviews with 25 of the 48 former *S*s and, without knowledge of the previous ratings, rated their needs at that time. Correlations between the later and the original ratings are .47 for ratings of n Ach and .42 for ratings of n Aff, both of which are significant at the .05 level. These figures may be interpreted either as indices of the "test-retest reliability" of the interview measure or as giving some empirical support to the conception of needs as being relatively stable over a period of time.

The two sets of sociometric ratings of their fellow group members by the 48 *S*s and their responses to the Postsession Questionnaire provide us with means to check upon the manipulation of the experimental conditions. These data (reported in Pepinsky, Hemphill, & Shevitz [14]) support the assumption that the *S*s' phenomenal view of the situation did correspond to the experimental intent. Separate chi-square tests by item show that in general under Condition A favorable sociometric choices were randomly distributed between the stooges and the *S*s; under Condition R, however, the stooges received a disproportionate share of negative choices.

The Postsession Questionnaire contained six items designed to test whether the *S*s felt relatively more personal acceptance under A and more rejection under R (e.g., "members seemed, a, to welcome, or, b, to resent my suggestions"). Chi-square tests by item of the association between conditions and the response categories are significant (at the .01 level or beyond in five instances; at the .05 level for the remaining item), and the trends are in the predicted direction. For each of two additional items, designed to determine whether the *S*s believed the falsified reports fed back to them, the responses show that nearly all *S*s under both conditions did believe the ratings reported to them: only three of the 48 *S*s thought that "the people in the group were not to be trusted"; four *S*s "felt that the ratings were completely false." Regardless of the *S*s' need classifications, their responses to the Postsession Questionnaire corroborate the effectiveness of the experimental strategy.

Separate intraclass correlations computed as checks upon interobserver agreement were .96 for the pairs of tallies of leadership attempts made during Session 1, .92 for Session 2, and .96 for the total tallies made during both sessions. Thus, these tallies provide a highly reliable measure of the dependent variable.

Test of Hypotheses

The data of major interest are the numbers of leadership acts attempted by the two classes of *S*s under Conditions A and R. Each measure is the sum of the two observers' tallies of the leadership acts attempted by each subject during one session. Since inspection revealed the presence of marked positive skewness, the raw data were transformed in order to obtain a more normal distribution and to stabilize the variance within treatments. This transformation was achieved by extracting the square root of the sum of the attempted leadership acts recorded by the two observers for each *S* during each session. The greatest integers smaller than these square roots are used as the individuals' "attempted leadership scores."

Application of the analysis of variance permits a detailed examination of the effects upon attempts to lead, not only of needs, but also of the other independent variables and their interactions. The results of this analysis are presented in Table 1. The F_1 column of the table reveals that the most conspicuous findings with reference to the major hypotheses are the failure of Needs and of the Conditions × Needs interaction to attain significance. It becomes evident, rather, that variation in the number of attempts to lead is definitely associated with the rejection-acceptance conditions. The effect of variation in the control of the conditions throughout the experiment, identified as Blocks (composed of two groups each), is significant at the .05 level. The interactions of Conditions × Blocks and of Conditions × Needs × Blocks are significant at the same level. The F_2 column shows the results obtained, when the significant lower order effects containing these components are retested against the Conditions × Needs × Blocks interaction. The effect of Conditions remains significant at the .01 level, whereas now the Blocks and Conditions × Blocks values do not meet the .05 criterion. We may conclude, therefore, from this more stringent test that the difference attributable to Conditions is not simply a product of variations between Blocks; individual differences in attempts to lead are dependent upon the special conditions of acceptance or rejection under which the *S*s worked on the task.

A point biserial correlation coefficient was computed to obtain an estimate of the extent of the association between the experimental conditions and attempted leadership scores. The coefficient is .61, which, as an underestimate of the true relationship, indicates that more than 36 per cent of the variance in attempts to lead may be accounted for by the conditions of acceptance or rejection to which the individual group members were subjected.

Group Productivity and Morale

Our records now permit us to examine the relationships between group productivity as measured by task success—i.e., amount of money earned—and four other variables: (*a*) attempted leadership scores; (*b*) experimental Conditions A and R; (*c*) judged quality of group decisions, where each requisition for model components countersigned by all group members is regarded as one such "decision"; and (*d*) group member "morale" as measured by individual stanine scores on two dimensions of the GDDQ—*Hedonic Tone* and *Viscidity*. In brief, *Hedonic Tone* is "the degree to which group membership is accompanied by a general feeling of pleasantness or unpleasantness"; *Viscidity* is "the degree to which members of the group function as a unit" without personal conflict and dissension (Hemphill, 7).

What we find first is that there is a pronounced positive relationship between the groups' profits and attempts to lead under Condition R, but that these variables are independent under Condition A. Kendall's tau was used to evaluate the correspondence between the group rankings arranged according to (*a*) amount of money earned and (*b*) order or magnitude of the summed attempted leadership scores of the two *S*s in each group. Under Condition R, tau is .84 (P less than .001), while under A tau is .00. Next, we find that although the average profit under R is \$1.80 as compared with \$3.37 under A, the apparent difference is not reliable ($t = 1.71$), due to wide variability present under both conditions.

The number of leadership attempts and the amount of group profit will be positively correlated only when suggestions are made and followed that consistently tend to be economically sound proposals. And yet, under Condition A, where no such relationship exists, net profit is at least as high as it is under Condition R, where money earned is directly associated with attempts to lead. These comparisons suggested that there might be a qualitative difference in the character of the groups' decisions under the two conditions, and so an analysis was made of the order forms submitted by the 24 groups. It was assumed that since each of these requisitions had to be signed by all the members of the group, each one represented group consensus about a plan of action, and they could properly be regarded as records of group decisions. Two judges independently assigned each completed form to one of two categories, "best" or "not-best" decisions. A "best" decision was defined as one that, if followed (i.e., if the toy for which specific parts were ordered was completed and sold), would result in the maximum profit to be realized at a particular time from the money enclosed with the form; and all other forms were placed in the "not-best" class. The judges agreed in their categorization of all 131 order forms. A test of the independence of the experimental conditions and the quality of group decisions yields a chi-square value of 6.55 (P less than .01) indicating that the two variables are significantly related. Departures of the observed cell values from expectancy are in the direction of fewer "not-best" decisions reached under Condition R and *more* decisions of this type reached under A. The A groups submitted more than twice as many orders as the R groups, or 88 as compared with 43 (a deviation from chance significance at the .01 level). But the latter groups turned in a disproportionately *small* number of orders representing "not-best" decisions. Sheer volume of orders is not a valid measure of productivity on this task, since higher activity level is not necessarily accompanied by significantly greater profit. Our impression is that under Condition R individuals were more self-critical and cautious about making suggestions than under Condition A, where they felt free to express any and all proposals that occurred to them.

When separate t tests were made of the mean differences in their stanine scores on Hedonic Tone and Viscidity, the means of the 24 *S*s run under A were significantly (P less than .01) higher than those of the 24 *S*s run under R. The correlations between profits and stanine scores (averaged to obtain a group measure for each of the 24 groups) were significant at the .05 level and positive for both dimensions: Hedonic Tone, .48, and Viscidity, .40. Without regard to experimental condition there was a hardly surprising tendency for making more profit to be accompanied by a general feeling of pleasantness and by less dissension.

DISCUSSION

To the extent that the operations performed did permit a test of the stated hypothesis, it must be rejected. When selected individuals were placed in a situation in which their attempts to lead resulted in personal rejection, those with a relatively high inferred need for achievement and a low need for affiliation were not able to lead more than those with relatively high affiliation and low achievement needs. We began with the question of who will be able to try to lead, if his attempts result in his own personal rejection. So far our best reply is, "Nobody," or "Well, *hardly* anybody," at least in the case of the "temporary, *ad hoc* group" (Strodtbeck, 20).

But despite the reported checks upon the experimental procedures, there still are grounds for doubt as to whether the experiment did permit an adequate test of the theoretical prediction. There are several alternative but not mutually exclusive possibilities: (*a*) the need measures were inadequate; (*b*) the two needs were not sufficiently differentiated in the samples employed; (*c*) an underlying characteristic of "other-directedness" (Riesman, 16) is so typical of the contemporary college population that otherwise accurately classified "high need achievers" are unable to withstand strong personal rebuff by their peers; (*d*) the required experimental control was not maintained so that there was—in the view of the *S*s—equivalent opportunity for the satisfaction of n Ach under both experimental conditions (cf. Asch, 1); and (or) (*e*) the observed reactions to such conditions would be different, if the groups involved continued to operate over a longer period of time.

On the positive side, the experiment is a clear-cut demonstration of the importance of situational factors as predictors of attempts to lead. The results are unambiguous in supporting the inference that, even after strenuous efforts have been made to select for study two classes of individuals whose behavior under "normal" circumstances indicates that they represent opposite "personality types," those persons are more alike than different, when exposed to strong and consistent personal rejection or acceptance. Such extreme conditions produce wide variability in the frequency of leadership attempts, and the effect of these treatments cuts across and obliterates individual differences in the measured personal characteristics. This outcome is consistent with the current emphasis upon situational variables as determinants of leadership behavior.

Our collateral findings pertaining to group productivity and morale call to mind Shaw's recent study (17) of group performance within various communication nets under appointed "authoritarian" and "non-authoritarian" leaders. He compares his results with those of Lewin, Lippitt, and White's (11, 12) earlier experiment where groups worked under trained "autocratic," "democratic," and "laissez faire" leaders. In both these cases, group climate was varied through the manipulation of the behavior of appointed leaders, whereas what "sound like" similar atmospheres were induced in the present experiment by different, but complementary procedures. Any generalizations about the effects of experimentally created climates upon group productivity must be qualified to the extent that—in contrast to what happens in natural group situations—the same conditions do not hold for *all* parties to the interaction. We have not been able, however, to resist the temptation to compare our own results with those of the other two studies on the provisional assumption that in a phenomenal sense the situations were comparable. If we take this inferential leap, some interesting parallels and contradictions appear.

These comparisons may be summarized in tabular form:

Authors	Experimental Conditions	Productivity		Morale
		Quantitative	Qualitative	
Lewin, Lippitt, & White	(*a*) Autocratic vs. (*b*) democratic leaders	Higher under (*a*) than (*b*)	Better under (*b*) than (*a*)	Higher under (*b*) than (*a*)
Shaw	(*a*) Authoritarian vs. (*b*) nonauthoritarian leaders	Higher under (*a*) than (*b*)	Better under (*a*) than (*b*)	Higher under (*b*) than (*a*)
Pepinsky, Hemphill, & Shevitz	(*a*) Rejection vs. (*b*) acceptance of potential leaders	No signif. difference between (*a*) and (*b*)	Better under (*a*) than (*b*)	Higher under (*b*) than (*a*)

When we take into account that the tasks assigned to the groups in the three studies were objectively quite different, that in Shaw's experiment the *S*s were limited to written communication, and that in the Lewin, Lippitt, and White study the *S*s were children working under adult leaders, it seems remarkable that there is as much apparent agreement as there is among the authors' conclusions. In all three experiments the morale of the *S*s was higher when they were in situations characterized by "warmth and permissiveness" as opposed to those characterized by "coldness and criticism." In the two earlier studies, however, productivity as measured by *quantitative* performance criteria was relatively high under autocratic or authoritarian leaders. In our study this kind of comparison shows no significant difference between conditions, but the absolute trend was in an apparently opposite direction. It may be legitimate to regard this result as a function of the manipulation of follower rather than leader behavior, but it is just as plausible to suspect that it is an effect of uncontrolled variation in the stooges' behavior. In two out of three studies a cold climate also seemed to produce *qualitatively* better performance. Lewin, Lippitt, and White say otherwise, but their statement seems to be based on informal judgment only.

In general, the results of these laboratory experiments are congruent with the repeated findings of studies made in field situations (Bradyfield & Crockett, 4) that "morale" and group "productivity" are seldom significantly and positively related, regardless of wide differences in the definition and measurement of these variables. These empirical data are, of course, offensive to the democratic *Zeitgeist.* But at least in the case of the experiment we have just reported, we can say that, although groups more frequently made "poor" decisions under acceptance conditions than under rejection conditions, they compensated for this by their greater activity so that in the long run they did at least as well in terms of money earned—and were happier besides!

SUMMARY

This report has described an experiment that was designed to test the hypothesis that under conditions such that to attempt to lead is to invite rejection, group members characterized by a relatively high need for achievement and a relatively low need for affiliation would make more frequent attempts to lead than would other members having relatively high affiliation and low achievement needs. The experimental design provided for a test of this hypothesis under controlled laboratory conditions and involved a total of 24 four-man groups.

The principal experimental findings are:

1. Under Condition R (rejection) individuals with high need Achievement did not attempt to lead with greater frequency than individuals with high need Affiliation. Therefore, the central hypothesis is not supported by the data.

2. Differences in attempts to lead are clearly attributable to differences in the experimental conditions to which the group were subjected, rather than to the major needs specified. There were significantly more attempts to lead under Condition A (acceptance) than under Condition R, regardless of individual differences in inferred needs.

3. A collateral analysis of group productivity, as measured by money earned, shows

that there was not a significant difference in net profits under the two conditions, but that fewer poor group decisions were made under R. Under Condition A, however, there was higher morale, and the groups compensated for their less consistently good decisions by maintaining a relatively high activity level so that, in the long run, they made as much money as the less sanguine rejection groups.

The results suggest that if we are to understand better the development of characteristic group atmospheres and the effects of such conditions upon behavior in natural groups, we do well to take into account the treatment of the leader *by* others, as well as his treatment *of* others.

REFERENCES

1. ASCH, S. E. Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men.* Pittsburgh: Carnegie Press, 1951. Pp. 177–190.
2. ASH, P. The reliability of psychiatric diagnoses. *J. abnorm. soc. Psychol.*, 1949, **44**, 272–277.
3. BASS, B. M., WURSTER, C. R., DOLL, PADDY ANN, & CLAIR, D. J. Situational and personality factors in leadership among sorority women. *Psychol. Monogr.*, 1953, **67**, No. 16 (Whole No. 366).
4. BRAYFIELD, A. H., & CROCKETT, W. H. Employee attitudes and performance. *Psychol. Bull.*, 1955, **52**, 396–424.
5. GUETZKOW, H., & BOWES, ANNE E. Preliminary report on the development of organizations with a division of labor. Pittsburgh: Carnegie Inst. of Technology, 1954. (Mimeo.)
6. HEMPHILL, J. K. A proposed theory of leadership in small groups. In J. K. Hemphill, Pauline N. Pepinsky, R. N. Shevitz, W. E. Jaynes, and Charlotte A. Christner, *Leadership acts, I. An investigation of the relation between possession of task relevant information and attempts to lead.* Columbus: Ohio State Univer. Res. Found., 1954. Appendix A.
7. HEMPHILL, J. K. Group dimensions: A manual for their measurement. *Ohio State Univer. Bur. Res. Monogr.*, No. 87. Columbus: Ohio State Univer., 1956.
8. HEMPHILL, J. K., PEPINSKY, PAULINE N., KAUFMAN, A. E., & LIPETZ, M. E. The effects upon attempts to lead of task motivation and expectancy of accomplishment of the task. *Psychol. Monogr.*, 1957, **71**, No. 22 (Whole No. 451).
9. HEMPHILL, J. K., PEPINSKY, PAULINE N., SHEVITZ, R. N., JAYNES, W. E., & CHRISTNER, CHARLOTTE, A. The relation between possession of task relevant information and attempts to lead. *Psychol. Monogr.*, 1956, **7**, No. 7 (Whole No. 414).
10. HUNT, W. A., ARNHOFF, F. N., & COTTON, G. W. Reliability, chance and fantasy in inter-judge agreement among clinicians. *J. clin. Psychol.*, 1954, **10**, 294–296.
11. LEWIN, K., LIPPITT, R., & WHITE, R. K. Patterns of aggressive behavior in experimentally created "social climates." *J. soc. Psychol.*, 1939, **10**, 271–299.
12. LIPPITT, R., & WHITE, R. K. The "social climate" of children's groups. In R. G. Barker, J. Kounin, & H. Wright (Eds.), *Child behavior and development.* New York: McGraw-Hill, 1943.
13. MCCLELLAND, D., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
14. PEPINSKY, PAULINE N., HEMPHILL, J. K., & SHEVITZ, R. N. *Leadership acts, II. The relation between needs for achievement and affiliation and attempts to lead under conditions of acceptance and rejection.* Columbus: Ohio State Univer. Res. Found., 1955.
15. RAFFERTY, JANET E. Use of two interpretive projective techniques for prediction within Social Learning Theory of Personality. Unpublished doctoral dissertation, Ohio State Univer., 1952.
16. RIESMAN, D. *The lonely crowd.* New Haven: Yale Univer. Press, 1950.
17. SHAW, M. E. A comparison of two types of leadership in various communication nets. *J. abnorm. soc. Psychol.*, 1955, **50**, 127–134.
18. SHEVITZ, R. N. *Leadership acts, IV. An investigation of the relation between exclusive possession of information and attempts to lead.* Columbus: Ohio State Univer. Res. Found., 1955.
19. SHIPLEY, T. E., JR., & VEROFF, J. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, **43**, 349–356.
20. STRODTBECK, F. L. The family as a three-person group. *Amer. sociol. Rev.*, 1954, **19**, 23–29.

Received June 17, 1957.

THE EFFECTS OF ANXIETY LEVEL AND PSYCHOLOGICAL STRESS ON VERBAL LEARNING

JANET A. TAYLOR
Northwestern University

USING the Manifest Anxiety Scale (MAS) as a selective device, a number of studies have been conducted by Spence, Taylor, and their associates in an attempt to test hypotheses derived from Hullian theory concerning the relationship between the performance of human *S*s in certain learning situations and level of total effective drive. Use of the MAS in this connection has rested on the assumption that the scores are related to level of emotionality and drive (D). Considerable experimental support has been found for these hypotheses (hereafter referred to, for convenience, as drive theory), thus confirming both the notion that MAS scores reflect level of D and the postulated interactions between drive level and learning task variables (10).

The major purpose of these studies has been, as indicated above, to investigate hypotheses concerning the role of D, the MAS having been developed and used merely as a convenient instrument for selecting *S*s differing in drive level. It is also possible, as many have preferred to do, to view the results of the studies comparing the performance of extreme scorers on the MAS as throwing some light on the personality pattern of manifest anxiety itself. There are undoubtedly many characteristics other than drive level on which such high anxiety (HA) and low anxiety (LA) *S*s differ; identification and incorporation of such characteristics into drive theory should broaden the usefulness of the latter and provide further understanding of anxiety as a personality variable.

One such characteristic, mentioned by several writers (1, 4, 6, 10) as possibly differentiating between the two extreme anxiety groups, concerns their reactions to psychological stress (e.g., reports of inadequate performance in an experimental task). The present writer, for example, has suggested that under conditions of psychological stress, internal responses of the type that Child (1) has referred to as task-irrelevant, are more easily or more intensely aroused in HA *S*s than in LA. In experimental tasks in which such extratask responses interfere with efficient performance (e.g., verbal learning), HA *S*s under stress conditions would therefore be expected to be inferior to neutral control *S*s, whereas the performance of LA *S*s should be affected to a lesser degree (10).

These hypotheses were suggested by the results of studies by Gordon and Berlyne (3) and by Lucas (4). In both of these investigations, HA *S*s told that their performance on a verbal learning task was inadequate were inferior in performance on a subsequent task to HA groups run under neutral conditions. The LA groups, in contrast, showed no decrement in performance under stress, when compared to their control groups. While the results of these studies seem to confirm the notion that HA *S*s are more liable to make interfering extratask responses under conditions of psychological stress, they could also be attributed simply to a greater emotional reaction to the stress instructions and hence a greater increase in drive level on the part of the HA groups. That is, the empirical predictions generated from drive theory state that increases in drive level facilitate performance in relatively simple tasks in which a single response tendency is evoked (e.g., classical conditioning); in more complex situations in which competing intratask responses are evoked and the correct response tendency is relatively weak, high drive (anxiety) *S*s tend to lose their superiority and, as the number of incorrect responses become greater and/or more dominant, to become inferior. Both the Gordon and Berlyne and the Lucas studies employed learning tasks of the competitional type. Thus it could be argued that the HA *S*s reacted with greater emotionality than LA groups to the stress conditions and the resultant increase in drive was responsible for their performance decrement.

The present study was designed to provide an experimental arrangement in which the effects of increasing drive levels would be expected to result in differences between HA and LA *S*s in the opposite direction to those expected if extratask, interfering responses were aroused by the stress condition. Specifically, verbal learning tasks were designed in such a manner as to minimize competing intratask responses (11). Hence, drive theory would predict that under neutral conditions, a HA (high drive) group would perform at a higher level than LA. If the introduction of stress results simply in an increase in drive level, and, further, HA *S*s are more reactive to such stress, these *S*s should increase their margin of superiority over the LA (when compared to neutral groups). If, on the other hand, the major effect of stress is to arouse competing extratask responses the HA group should no longer exhibit a performance superior to the LA and may even be inferior to them.

METHOD

Subjects. The high anxiety (HA) group consisted of 40 *S*s who scored 23 or above on the MAS and the low anxiety (LA) group of 40 *S*s scoring 9 or below. Each group was further subdivided into a stress and a neutral group of 20 *S*s each, individuals being alternately assigned to the subgroups as they appeared to be tested. All *S*s were drawn from introductory psychology classes to whom the MAS had been given and were naive both with respect to previous experience in verbal learning studies and the purpose of the experiment.

Materials. A practice list and two equivalent experimental lists were employed, each consisting of eight pairs of nonsense syllables. The practice list contained syllables of 47–53 per cent association value (2) and the experimental lists syllables of 0–20 per cent association value. Both interlist and intralist similarity were at a minimum. The specific lists employed were ones that had been used in a previous study (11) in which it had been predicted and found that HA *S*s perform at a level superior to LA *S*s.

The lists were presented by means of a Hull-type memory drum at a 2-2 sec. rate with 4 sec. between trials. The syllables were typed on an endless white tape, three orders being prepared for each list to minimize serial learning.

Procedure. Following preliminary instructions, which included the statement that performance on learning tasks is related to intelligence, each *S* was given the practice list for 10 trials and then each of the two experimental lists for 15 trials. Order of presentation of the two experimental lists was counterbalanced within each of the four subgroups. For *S*s in the neutral, nonstress subgroups, the instructions given before each of the experimental lists indicated merely that *S* was to learn another list of the same type. Before receiving the second experimental list, however, *S*s assigned to the stress condition were told that their performance on the previous lists had been considerably poorer than most *S*s, asked whether they had been trying hard or felt well and told that it was hoped they would be able to improve their record. At the end of the experiment, all stress *S*s were told of the deception, assured that their performance had been adequate, and cautioned not to discuss the nature of the experiment with others.

TABLE 1

MEAN NUMBER OF TOTAL CORRECT RESPONSES OF THE HA AND LA GROUPS ON THE FIRST EXPERIMENTAL LIST

	HA	LA
M	40.98	33.78
σ	16.35	14.81

RESULTS

On the first experimental list, given to all *S*s under neutral conditions, the HA group exhibited the superiority in performance predicted by drive theory, as may be seen in Table 1. The total number of correct responses given by each *S* were analyzed statistically by a 2 x 2 analysis of variance summarized in Table 2, which permitted a test of the equivalence of the subgroups as well as of the effects of anxiety level. The *F* for the HA-LA comparison was significant ($p < .05$), as anticipated. The terms for conditions and for the interaction between groups and conditions were not significant, thus suggesting that within each anxiety group, the subgroups were adequately matched.

Perhaps the clearest demonstration of the effects of the stress condition on second list performance is afforded by obtaining a difference score for each *S* (total number correct responses on List 2 minus total number on List 1). The mean difference scores for each of the

TABLE 2

ANALYSES OF VARIANCE OF NUMBER OF TOTAL CORRECT RESPONSES ON THE TWO EXPERIMENTAL LISTS

Source	List 1				List 2			
	SS	*df*	MS	*F**	SS	*df*	MS	*F**
Groups	1036.45	1	1036.45	4.11	1224.49	1	1224.49	3.88
Conditions	288.45	1	288.45	1.14	3137.51	1	3137.51	9.95
G × C	.70	1	.70	—	70.44	1	70.44	—
Within	19187.90	76	252.47		23973.05	76	315.44	

* For 1 and 70 *df*, $F = 3.98$ at .05 level and at. 701 at .01 level

TABLE 3

MEAN DIFFERENCE BETWEEN THE TWO EXPERIMENTAL LISTS (LIST 2 – LIST 1) FOR THE FOUR SUBGROUPS

	HA		LA	
	Stress	Neutral	Stress	Neutral
M	−5.10	+5.25	−4.10	+3.00
σ	14.87	12.36	16.93	9.66

four subgroups are reported in Table 3. While both of the subgroups tested under neutral conditions improved in performance on the second list (presumably because of practice effects), the two stress subgroups were poorer.

An analysis of variance of the second list data, shown in Table 2, indicated that the *F* for conditions was significant at the .01 level while the *F* for the anxiety groups was significant between the .05 and .10 levels. The *F* for the interaction between groups and conditions, however, was less than 1.00. Thus, although the stress condition had the effect of interfering with performance, the HA *S*s did not show any greater responsiveness to it and within each experimental condition, maintained their previous margin of superiority over the LA.

DISCUSSION

To summarize the results of the present study, HA *S*s under the neutral, nonstress conditions performed at a superior level to the LA group, as was predicted by drive theory. The subgroups told that their performance had been inadequate showed a significant decrement in subsequent performance when compared to their neutral controls. If it is assumed that the superior performance of the HA *S*s tested under neutral conditions was due to their higher drive level, the poorer performance of the stress groups can be taken as indicating that the major effect of the stress instructions was to arouse extratask irrelevant responses which interfered with efficient performance. No evidence was found, however, to support the contention that HA *S*s are more prone to exhibit these extratask responses, i.e., there was no interaction between anxiety level and the stress-neutral conditions. Superficially, these results appear to contradict the previous studies of Gordon and Berlyne (3) and of Lucas (4), both of which found a greater inferiority of HA groups under stress conditions than under neutral conditions. That is, whether one wished to interpret their results as due to the greater emotional responsiveness (and hence proportionately higher drive level) of HA *S*s to the stress or to a greater tendency to make irrelevant responses under stress, some type of differential reaction to stress by the HA and LA groups might have been anticipated in the present study. One possible reconciliation is to assume that in HA *S*s both drive level and extratask responses increase proportionately more in stress situations than in LA; in the present situation (employing materials with low intratask competition) these two factors affect performance in opposite directions, increased drive tending to facilitate and irrelevant responses to interfere. Thus the two may have cancelled each other. In the previous studies, on the other hand, the learning tasks were such that both factors could be assumed to contribute to the performance decrement, thus leading to the greater apparent susceptibility of the HA groups to stress.

The results of the present study also have bearing on several alternative hypotheses that have been offered in place of drive theory to explain differences in performance that have been found between HA and LA groups. The hypothesis that HA *S*s react to stress situations with irrelevant responses to a greater degree than LA was originally proposed by Child (1). Child, however, applied this notion to all of the studies carried out to evaluate drive theory, including those in which no attempt was made to introduce psychological stress. In common with proponents of drive theory, Child also assumed that HA *S*s react more emotionally and hence have a higher drive level in experimental situations than LA. He stated further that for the HA individual, this emotionality always serves as a cue for the evocation of strong irrelevant extratask responses, responses which are weak or relatively lacking in LA *S*s. In simple types of situations (e.g., classical conditioning) in which performance of the correct response is relatively invulnerable to the disruptive effects of irrelevant responses, HA *S*s should be superior due to their higher drive level. The performance of relatively complex tasks (e.g., verbal learning), however, can be disrupted by extratask responses. The extratask responses of HA *S*s interfere with performance more than their higher drive level facilitates it, thus resulting in the performance of HA *S*s being inferior to LA groups in these situations, even under so-called "neutral" conditions. The addition of stress instructions serves to increase the drive level and, hence, the number and/or intensity of irrelevant responses in HA *S*s, and therefore magnifies their inferiority.

The data available at the time Child offered his hypotheses concerning the correlation between drive level and extratask responses in HA *S*s seemed to support his contentions about as well as they did drive theory, which emphasized the importance of intratask response competition in determining the relative performance of HA and LA groups. However, several later investigations of verbal learning (5, 8, 9, 11) have demonstrated the superiority of HA *S*s in tasks with minimal intralist interference. As the writer has previously pointed out (10), these results, while predicted by drive theory, seem to weaken the plausibility of Child's rival hypothesis since it would appear to be difficult to maintain that verbal learning tasks of low intralist competitiveness are not susceptible to extratask interference. The design of the present study, however, appears to provide a direct test of Child's hypotheses. That is, a superior performance of HA *S*s, as compared to LA, would indicate, according to Child's notions, that the task employed must have been of the type in which performance is relatively unaffected by the presence of irrelevant responses. If stress were introduced into a situation employing such a task, the drive level of HA *S*s (and perhaps of the LA as well although to a lesser degree) and hence the magnitude of irrelevant responses would be raised. But, since the latter cannot interfere with performance on this type of task, the HA group, with their increased drive, should show even greater superiority of performance when compared to LA *S*s than when operating under neutral conditions. Under the neutral conditions of the present study, HA *S*s were superior to the LA, indicating that the lists, according to Child, were relatively invulnerable to extratask responses. For the stress condition, Child would therefore predict an increased margin of superiority of the HA group over the LA, as well as the superiority of the HA stress *S*s over the HA neutral group. The actual results, of course, showed the opposite: a *decrease* in performance in both stress groups and to the same extent in the LA and HA groups. Thus Child's hypotheses concerning the crucial role of irrelevant response tendencies were not upheld.

In a recent study, Saltz and Hoehn (6) also proposed an alternative to drive theory based in part on differential reactions of HA and LA groups to stress. They suggest that previously reported differences between HA and LA groups (under neutral conditions) are not related to the interaction between drive level and the number and relative strengths of competing intratask responses as drive theory has postulated but rather to the "difficulty" of the tasks. As task difficulty (which, they state, had previously been confounded with task complexity) increases, the anxiety level of HA *S*s rises proportionately more than that of LA *S*s, and this anxiety, in some manner which they did not specify, leads to response decrement.

These speculations, insofar as they are susceptible to empirical evaluation, do not appear to be supported by available evidence. The lack of differential reaction of the anxiety groups to stress in the present study casts doubt, of course, on the special role of this variable. With respect to the effect of difficulty, their statements imply that in "easy" tasks, HA *S*s should be *equal* in performance to LA and become increasingly inferior as tasks become difficult. A number of investigations (10) involving both nonverbal and verbal tasks (including those of the present study) have, however, shown HA groups to be *superior* in performance. Further, examination of learning studies does not show, in as far as it is possible to make judgments of relative difficulty, any consistent relationship between the "ease" or "difficulty" of a task and the direction of performance difference between HA and LA groups. For example, the experimental lists of the present study consisted of 8 (noncompetitional) pairs of nonsense syllables of 0–20 per cent association value in contrast to 15 (noncompetitional) pairs of adjectives employed in an investigation by Spence, Farber, and McFann (8). In spite of its greater length, the adjective list was considerably "easier" than the first experimental list of the present study (e.g., at trial 10 for the total groups, a mean of 3.3 correct responses out of

8 syllables vs. a mean of approximately 14.3 out of 15 adjectives). Although there were marked differences in difficulty, HA *S*s were superior on both lists. Spence, Farber, and McFann also reported the results of a second adjective list, involving both competitional and noncompetitional pairs. This list, too, was "easier" than the syllable list of the present study, yet HA *S*s were superior to the LA on the noncompetitional pairs and *inferior* on the competitional pairs.

In their investigation, Saltz and Hoehn manipulated, variously, the association value and formal intralist similarity of serial lists of nonsense syllables in an attempt to obtain lists of varying difficulty level and degree of intralist competition. Their results, they stated, were contradictory to drive theory and supportive of their difficulty notions. A number of features of the study, however, make their findings difficult to interpret. The types of lists employed and their rules of construction make predictions in terms of drive theory troublesome if not impossible. As Spence (7) has pointed out, in order to make predictions about any learning situation from drive theory it is also necessary to have a specific theory about the variables involved in the learning task. Neither Spence nor the writer has presented a theory of serial learning, let alone any hypotheses as to the role of association value of nonsense syllables, nor did Saltz and Hoehn.

Quite independent of the adequacy of the lists for testing drive theory, a closer examination of their results does not supply the support for the difficulty hypothesis that the writers assert. Saltz and Hoehn employed three serial lists of equal length, specifying that the relative difficulty of each be defined in terms of the performance (number of trials to criterion) of the LA *S*s. Two of the lists (of competing 90 per cent association value material and of noncompeting 13 per cent association material) were chosen to be equal in difficulty (i.e., no significant difference in the performance of LA groups). On both of these lists, there was no significant difference in the performance of the HA and LA groups.[1] On the third list consisting of noncompetitional 0% association value material the HA were significantly inferior to the LA, the approximate difference in mean number of trials to criterion being a striking 13 trials. This inferiority constituted the writers' evidence for support of their difficulty hypothesis. Their conclusion rests on the stated assumption, that was not tested, that the zero per cent list was more difficult than the 13 per cent list (on which LA and HA did not differ). When the performance on the zero per cent and 13 per cent lists *are* compared for the LA groups, whose performance, it will be recalled, had been designated as defining the relative difficulty of the tasks, practically identical mean trials to criterion can be noted. The general literature on the effects of zero vs. 13 per cent association value does not make this equivalence surprising but it does fail to confirm the critical assumption that their zero list was more difficult. Thus, we are faced with what reasonably might be called three lists of equal difficulty, the HA *S*s being equal in performance to the LA on two of them and inferior by a disconcertingly large margin on the third. This inconsistency might be attributed to the writers' choice of military personnel (airmen) as *S*s. Correlations between measures of intellectual ability and MAS scores have often been found with such intellectually heterogeneous groups. While in part such correlations might be attributed to the disruptive effects of high anxiety (drive?) level on the performance of complex tasks, there is good evidence (12) to suggest that genuine intellectual differences are represented among high and low scoring groups drawn from such intellectually heterogeneous populations due to the greater ability and/or willingness of the more intelligent to recognize the purpose of the scale and distort their answers in a "healthy" direction. Thus, their series of groups may have been contaminated in unknown degrees by shifting intelligence levels, "false negatives" in the low scoring groups, etc. In short, previous investigations do not support the difficulty hypothesis, while the

---

[1] Saltz and Hoehn made no prediction about whether or not any differences should be expected between HA and LA *S*s (although it might be pointed out that in terms of the type of materials selected and the mean number of trials to criterion taken by the groups, the lists were, in usual terms of reference, on the difficult side and some inferiority of the HA *S*s might have been expected). The only "confirmation" of the difficulty hypothesis in this comparison lay in the rather negative finding that with two lists of equal difficulty, the lack of significance found between the HA and LA groups in one list was also found in the other.

Saltz and Hoehn study itself seems inadequate to provide any meaningful test.

SUMMARY

On the assumption that scores on the MAS reflect drive level, a number of studies have been conducted with high and low anxiety groups to test hypotheses concerning the interaction between drive level and certain learning task variables in determining performance. The present study attempted to investigate the suggestion that another characteristic on which high and low anxiety *S*s differ is their susceptibility to irrelevant, extratask responses under conditions of psychological stress. It was predicted that under neutral conditions high anxiety (high drive) *S*s would exhibit a performance superior to that of low anxiety (low drive) *S*s on a paired-associate learning task with minimal intratask interference but that under conditions of psychological stress (report of inadequate prior performance) high anxiety *S*s, due to the greater arousal of interfering extratask responses, would no longer exhibit the superiority found under neutral conditions. Results indicated that while the high anxiety *S*s under neutral instructions were significantly superior to the low anxious, as predicted, and the *S*s operating under stress were inferior to their neutral controls, the predicted interaction between anxiety level and stress was not found. The implications of these results for several rival hypotheses of the drive interpretation of anxiety were discussed.

REFERENCES

1. CHILD, I. L. Personality. *Annual Rev. Psychol.*, 1954, **5**, 149–170.
2. GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.
3. GORDON, W. M., & BERLYNE, D. E. Drive level and flexibility in paired-associate nonsense-syllable learning. *Quart. J. exp. Psychol.*, 1954, **6**, 181–185.
4. LUCAS, J. D. The interactive effects of anxiety, failure, and interserial duplication. *Amer. J. Psychol.*, 1952, **65**, 59–66.
5. RAMOND, C. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120–124.
6. SALTZ, E., & HOEHN, A. J. A test of the Taylor-Spence theory of anxiety. *J. abnorm. soc. Psychol.*, 1957, **54**, 114–117.
7. SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.
8. SPENCE, K. S., FARBER, I. E., & MCFANN, H. H. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 296–305.
9. SPENCE, K. W., TAYLOR, J., & KETCHEL, RHODA. Anxiety (drive) level and degree of competition in paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 306–310.
10. TAYLOR, JANET A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.
11. TAYLOR, JANET A., & CHAPMAN, JEAN P. Paired-associate learning as related to anxiety. *Amer. J. Psychol.*, 1955, **68**, 671.
12. VOAS, R. B. Intelligence and the distortion of responses on the Taylor anxiety Scale. *Psychol. Rep.*, 1956, **2**, 87–89.

Received June 17, 1957.

PATTERNS OF MANAGERIAL TRAITS AND GROUP EFFECTIVENESS

EDWIN E. GHISELLI AND THOMAS M. LODAHL
University of California

In some group task situations the final product is merely the sum of the efforts of the individual members. In others, the productive end requires an interaction among the members and therefore their efforts must be coordinated. Particularly with the second type of situation, a substantial portion of the group's attention must be devoted to planning and the integration and direction of the activities of the individual members. Planning, integration, and direction may be termed managerial functions inasmuch as they pertain to the organization of the operations of the individuals so that they are performed in a harmonious and effective manner with respect to achievement of the group goal.

We are defining managerial functions, then, as those which are concerned with the organization, government, and control of the activities of the members of the group. These functions involve such activities as planning, policy making, execution, supervision, and decision making. Consequently, managerial traits would be those psychological characteristics of the individual that directly pertain to these functions. It can be argued that all characteristics of the individual are included within this definition of managerial traits. Thus, intelligence might be termed a managerial trait since it is related to effectiveness in planning and judgment. However, we do not here wish to make any rigorous classification of traits but merely to distinguish a class of them which in terms of the operations involved seem in one way or another to be intimately associated with managerial functions. The investigation of such traits would seem to be especially helpful in seeking to understand the contribution of individuals to the group endeavor.

There are two managerial traits which seem to us to be of particular significance in terms of the operation of the group as an organization. One trait is that of supervisory ability. It is apparent that there are a variety of ways in which supervision can be accomplished and that effectiveness of supervision is determined by a variety of situational factors such as the group's culture, the physical facilities, and the group task goal. The effectiveness of a supervisor can be gauged in three ways. First, the product of the group he supervises can be compared with that achieved by the same or similar groups working with different supervisors. Secondly, it can be evaluated in terms of the perception of his effectiveness on the part of those he supervises, as by subordinates' ratings of their superiors. Finally, a supervisor's effectiveness can be gauged by how he is perceived by his superiors, as by means of superior's ratings of their subordinate supervisors. While all of these ways of evaluating the effectiveness of supervision provide pertinent information, the latter would appear to have special significance with respect to organizational problems since it is accomplished by the governing persons. It is with supervisory ability as evaluated in this way that we wish to concern ourselves in the present investigation.

The second trait is concerned with what might be termed the decision-making approach. At one extreme there are those individuals who with confidence in their own individual talents and resourcefulness have the capacity to make decisions on their own. At the other extreme are those individuals who are careful planners and who feel the need to act with caution and to examine all the facts before arriving at a decision. We believe that this particular dimension of decision-making approach should be investigated in relation to group functioning since in industrial situations it differentiates between top management personnel, that is, those who give general direction to the organization, and middle management, those who carry out general policy and give it specific implementation (2).

The purpose of the present investigation is to examine the relationship between the distribution of the traits of supervisory ability and approach to decision making, as here defined, in relation to the organizational characteristics and performance of small groups in a task situation requiring continuous interaction and coordination among the members. Specifically, with respect to both traits we wish to examine the importance of (*a*) the average amount of the trait as possessed by the members of a

TABLE 1

RELIABILITY COEFFICIENTS OF AND INTERCORRELATIONS AMONG THE INDICES OF LEARNING

Index	1	2	3	4	5	6	7	8
1. Total Trips	.942	.007	.928	.052	−.556	−.370	.299	.771
2. Total Wrecks		.595	.005	.594	.698	.501	−.017	.093
3. Trips Last 4 Trials			.933	.072	−.474	−.364	.414	.895
4. Wrecks Last 4 Trials				.599	.490	.842	−.186	.033
5. Wrecks/Trips, Total					.736	.726	−.367	−.360
6. Wrecks/Trips, Last 4 Trials						.404	−.601	−.318
7. Correlation Trips and Trials								.561
8. Rate of Learning, Trips								

group, (*b*) the presence in the group of a person standing high in the trait, (*c*) the presence of a person who relative to the other members of the group is uncontested in the amount of the trait he possesses, and (*d*) the presence of such an uncontested person when the remainder of the group is homogeneous in the trait.

METHOD

After exploring several possibilities for a task situation, it was decided that a small model railroad that required multiple controls to operate was most appropriate. The task was complex, involving both intellectual and motor components, and required close cooperation among all members of the group. There were two sets of identical control panels; half of each set contained controls for siding switches, and the other half, the power controls for the various sections of track. Up to four people could operate the railroad simultaneously. There were two trains on the track, and the task was to run both trains around the track in opposite directions as many times as possible in a three-minute trial. The group was given two points for each complete circuit of the track, with the proviso that both trains must make the same number of trips in any one trial. In order to minimize recklessness, the group was penalized five points for each wreck or derailment. There were 12 three-minute trials, with a one-minute rest period after each trial. The *S*s were allowed to talk to each other at any time.

The *S*s were males, drawn from courses in psychology. Included in this analysis are data on 90 *S*s, 10 groups each of 2, 3, and 4 people.

Measurement of Experimental Variables

Performance in a learning task can be measured in a variety of ways. It is possible to take cognizance of correct responses and errors, total performance during the practice period and rate of improvement, final level of proficiency and consistency of performance. In the model railroad situation used in the present investigation, eight different indices were obtained: number of total trips, total number of wrecks, number of trips during the last four trials (as a measure of final level of performance), number of wrecks during the last four trials, the ratio of wrecks to trips during the entire practice period (as a measure of quality of performance), the ratio of wrecks to trips during the last four trials, the coefficient of correlation between trips by trial and number of trial (as a measure of consistency of performance), and the slope of the learning curve of trips (as a measure of rate of improvement). The best fitting straight line was taken as an indication of slope. Wrecks declined only slightly during the practice period and hence slopes were not determined for them.

The eight measures of performance first varied with the size of the group. In general the larger the group the better the performance as measured by indices involving trips and the poorer as measured by indices involving wrecks. Furthermore, the distribution of scores on all indices, particularly the last six, were skewed. To adjust for differences in performance resulting from group size and to correct for skewness, the scores on each of the eight indices within each sized group were transmuted to normal standard scores.

For the first six indices, reliability coefficients were determined from the coefficient of correlation between group normalized scores on odd and even trials adjusted by the Spearman-Brown formula. These coefficients are shown in the diagonals of Table 1. Two of these reliability coefficients are quite high, one reasonably satisfactory, two borderline, and one a bit low. As a whole these various indices of group performance display moderately satisfactory precision describing the performance of the groups.

In order to examine the dimensionality of these indices the intercorrelations among the eight scores earned by the groups were calculated. These coefficients are shown in Table 1. Following Tryon's (3) method, the intercorrelations among the eight measures were subjected to a cluster analysis. The results indicated that there are two principal clusters as shown in Table 2. One cluster includes all of the measures involving trips, and the other cluster the measures involving wrecks and quality of performance. It would seem reasonable to label the first cluster productivity and the second cluster recklessness, or, if the direction of the dimension is reversed, carefulness.

Table 2 shows that the measures that define each of the clusters are substantially correlated with the domain. If the four measures defining the first cluster are taken as a sample of the domain, the scores on the sample are estimated to correlate with domain scores .938. Taking the four measures defining the second cluster as a sample, scores on the sample are estimated to correlate with domain scores .937. Therefore, both samples are quite good indicators of their domains. The

TABLE 2

COEFFICIENTS OF CORRELATION BETWEEN MEASURES AND CLUSTER DOMAINS

	Cluster Domain	
Measures Defining Cluster I	I	II
Total Trips	*.829*	−.271
Trips Last 4 Trials	*.929*	−.205
Correlation Trips and Trials	*.529*	−.366
Rate of Learning, Trips	*.924*	−.172
Measures Defining Cluster II		
Total Wrecks	.027	*.747*
Wrecks Last 4 Trials	−.009	*.801*
Wrecks/Trips, Total	−.547	*.797*
Wrecks/Trips, Last 4 Trials	−.514	*.861*

coefficient of correlation between scores on the two domains, production and carefulness, is estimated to be .319. Thus, while scores on the two domains are not completely independent their relatedness is low. There appears to be a slight tendency for groups who are relatively more productive to be relatively more careful.

A score for each group was determined for both clusters of performance variables by summing the four scores on the defining variables. All variables were equally weighted except for trips and trials which, because of its substantially lower domain validity, was weighted only half. In the following analysis these two domain scores, productivity and carefulness, will be related to scores on the inventory used for measuring supervisory ability and approach to decision making.

From the observational data recorded by the experimenter while the group was performing the task, it was possible to obtain two more scores for each group, the number of changes in organization and the number of changes in method. An organizational change was defined by (*a*) any trading of tasks from one member to another, or (*b*) any change in the way in which the work was divided. An example of the first type would be if *S*s A and B were operating block controls and C operated switches, then C trades with B, so that now C operates blocks and B operates switches. An example of the second type would be if A and B are operating block controls and C operates switches, then B stops operating blocks and begins to direct the operations of A and C.

There were two main ways to approach the assigned task of getting the trains around the track. One was to run both trains simultaneously, using the sidings to pass, and the other was to run only one train at a time either for one trip or several, stopping it at a siding, then running the other train. The first was termed the simultaneous solution, the second, the successive solution. A change in methods was scored whenever a group changed from one method to the other. Split-half reliability coefficients were computed for these two types of changes, between odd and even trials. The corrected coefficient of reliability for organizational changes was .828, and for methods changes, .690.

Measurement of Predictor Variables

Ghiselli (1) has devised a Self-Description Inventory consisting of 64 personally descriptive adjectives presented in forced choices. This inventory yields several scales, one of which is a scale of supervisory ability. This scale was validated by comparing the scores earned by persons holding supervisory positions in lower and middle management positions with ratings of their effectiveness as supervisors made by their superiors. The scale, then, would seem to measure the type of supervisory ability needed for the present investigation.

Porter and Ghiselli (2) compared the responses to the items of the Self-Description Inventory made by persons in top management and in middle management. Examination of the adjectives checked by each group revealed that people holding top management positions tended to perceive themselves as being active and self-reliant, and willing to take action on the basis of confidence in themselves and their own abilities. In social relations they are equally confident and see themselves as straightforward and dignified. People holding middle management positions tend to see themselves as careful in planning and thoughtful in actions, seldom making rash decisions. They seem to wish to avoid giving the appearance of being controversial personalities and of exhibiting self-centered behavior. Taking the 21 items on the inventory that Porter and Ghiselli found to differentiate top from middle management personnel and give the foregoing self perceptions, a scale was formed which we have here termed Decision-Making Approach (DMA). This scale probably measures more than just approach to decision making as we have defined it here. However, it is the best available to measure the quality in which we are interested and at least reflects it as indicated by the individual's self perceptions.

For the *S*s used in the present investigation, the coefficient of correlation between scores on the two scales was .365. Therefore, it is apparent that while they measure certain common traits, the overlap is not great. In 20 of the 30 groups the same individual earned the highest scores on the two scales.

In order to examine the problems outlined earlier, three scores were computed for each of the 30 groups on the two scales, supervisory ability and decision making. First of all, for each group the scores of all members were averaged. These mean scores give the general level of the group on the traits measured. Secondly, in each group the highest score earned by any of the members was noted. Finally, the score earned by the second highest person was noted and the difference between it and the highest score was taken as an index of the extent to which the highest scoring member in each group was uncontested.

Finally, since we wished to examine the importance of a group having an uncontested person with the remainder of the group being homogeneous in the trait, a separate score was developed for the twenty groups containing three and four members. Two indices entered into this score, the first being the difference score just described, and the second being the range of scores of the two members of three-person groups and the three members of the four-person groups earning the lowest scores. For this purpose both difference and range scores were transmuted into standard scores and for each group the latter was subtracted from the former. As an example of a group earning a high composite score, the highest scoring individual might have a raw score of 28 on the DMA scale; the next highest man in the group might have a raw score of 16. This difference score would then be 12, corresponding to a standard difference score of 65. The lowest three scores in the group might be 11, 13, and 16. The range of these three numbers is 5, corresponding to a standardized range score of 35. The composite score then is 65 minus 35, or 30. As an example of a group with a low composite score, the highest raw score might be 20 and the next highest 19; the difference score here is 1, corresponding to a standard difference score of 30. The lowest three scores might be 12, 17, and 19; the range of these three scores is 7, corresponding to a standard range score of 45. The composite score is then 30 minus 45, or minus 15. It will be noted that the mean amount of the trait possessed by the group is the same in both cases, namely, 17. The first case represents substantial positive skewness, the second, substantial negative skewness. Since the numbers of cases in these groups is so small, it was not felt justified to use precise measures of skewness such as the third moment. Nevertheless, for purposes of communication this composite score has been termed degree of positive skewness.

RESULTS

The coefficients of correlation between the various group scores and the two measures of performance, productivity and carefulness, are shown in Table 3. The coefficients of correlation that are reliability different from zero are indicated by means of asterisks.

For the supervisory scale, there are no significant relationships between either of the measures of group performance and mean scores of the entire group or the highest score in the group. These results mean that whether a group does or does not possess a person who stands high in the trait are not significant factors in its performance.

TABLE 3

COEFFICIENTS OF CORRELATION BETWEEN THE VARIOUS GROUP SCORES AND THE PRODUCTIVITY AND CAREFULNESS CRITERIA

Type of Score	Productivity	Carefulness
Supervisory Scale		
Mean score of the entire group	.071	−.221
Highest score in the group	.151	−.327
Difference between the two highest scores	.436**	−.316
Positive skewness of group scores[a]	.454*	−.238
Approach to Decision Making Scale		
Mean score of the entire group	.165	−.362*
Highest score in the group	.265	−.237
Difference between the two highest scores	.435**	−.130
Positive skewness of group scores[a]	.815***	.249

* $p < .05$.
** $p < .02$.
*** $p < .001$.
[a] For groups containing 3 and 4 members only.

The results also indicate that the greater tendency for a group to have an individual who is uncontested in supervisory ability, the greater is the productivity of the group. However, the relative homogeneity or heterogeneity of the other members of the group apparently is of little importance since the coefficient of correlation for difference scores and of positive skewness scores, which include measures of both the difference score *and* relative homogeneity, are very much the same.

While none of the coefficients of correlation between any of the supervisory scale scores are related to carefulness of performance with usually accepted levels of statistical significance, it is interesting to note that they are all negative. This finding suggests that supervisory ability as gauged by the governing bodies of organizations is oriented toward production but not toward quality of performance.

The mean scores earned by the groups on the DMA scale are not significantly related to productivity but are to carefulness. By and large, those groups containing persons who tend to the dynamic type of decision making are less likely to be careful than those containing persons who tend to decide only after careful examination of the situation.

Whether or not the group contains a pers who stands high in approach to decision making, that is, tends to be self-sufficient in decision making, is not significantly related to group performance. However, if the group possesses such a person and his position in the group on this trait is relatively uncontested, then the group is likely to be superior in productivity. This relationship has statistical significance.

When the group possesses an individual who is uncontested in this trait and the remaining members of the group are homogeneous with respect to it, then there is a very high degree of likelihood that the productivity of the group will be high. While the relationship with the criterion of carefulness is not statistically significant it is interesting to note that it, too, is positive. In fact, this is the only positive relationship between any group score and carefulness.

Fig. 1 shows the relationships between the DMA positive skewness score, the numbers of

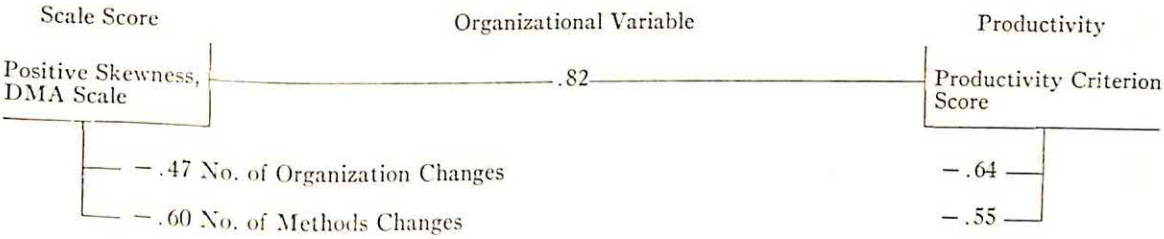
Scale Score
Organizational Variable
Productivity
Positive Skewness, DMA Scale
.82
Productivity Criterion Score
−.47 No. of Organization Changes
−.60 No. of Methods Changes
−.64
−.55

FIG. 1. RELATIONSHIPS BETWEEN SCALE SCORES, ORGANIZATIONAL VARIABLES, AND PRODUCTIVITY

organizational and methods changes, and the productivity criterion. The numbers of changes are negatively related both to the criterion and to the DMA positive skewness score. This result suggests that a large number of these changes leads to poor productivity, and that the groups with high positive skewness on the DMA scale tend to change organization and methods less frequently than negatively skew groups. All correlations shown in Fig. 1 are significantly different from zero beyond the .01 level of confidence.

DISCUSSION

There seem to be two dimensions in the task of operating the model railroad used in this experiment, namely, productivity and carefulness. Of the two, the predictor measures used were related much more closely to productivity than to carefulness. Furthermore, it appears that neither group means nor the highest score in the group on the scales used are very effective predictors of productivity. Measures of the skewness of groups' scores seem to be highly related to productivity, more so for the decision-making approach (DMA) scale than for the supervisory scale. In brief, it seems that the characteristics of the distribution of these managerial traits in the group are more important in determining its performance than any one measure of central tendency or extremity. Possible foundations of these findings need to be examined.

It seems to the authors that since the DMA skewness score describes a particular characteristic of the distribution of traits in the group considered as a whole, one place to look for a basis for its relationship to productivity might be in the organizational behavior revealed by the group at work. Fig. 1 suggests a promising approach to this problem. By and large, groups high on the DMA scale (positive skewness) went through far fewer organizational and methods changes than those which were low (negative skewness), and the number of these changes in turn is related to productivity.

It is not difficult to imagine why the number of organization and methods changes affects productivity. Each time a change occurs, some or all members of the group must learn a new set of responses to a new set of relevant stimuli. They must learn what new things to look for and what to do about them. Where there is a division of labor in the group, and this was necessary in order to perform the task, each change of organization or method involves such relearning, and the more changes there are, of course, the less opportunity the group has to learn any one method or set of tasks thoroughly. As to why the DMA positive skewness scale is related to the number of changes, only speculation is possible at this time. It may be that in groups with high positive skewness, the flow of suggestions coming from the "idea man" can be effectively put into practice or rejected by the two or three others who are "cautious evaluators." In groups low or negatively skewed, on the other hand, there may be a great many suggestions which are ineffectively evaluated, thus leading to many changes in organization. An analysis of the verbal interaction of the group might go far in answering some of these questions.

It must be emphasized that the sampling generality of these results is limited. The population sampled was undergraduate college males. They had never seen each other before coming into the experimental room, nor had they ever performed this particular task. A separate analysis by the authors[1] showed that a great deal of learning took place during the two hours spent on the task. Examination of the number of organization and methods changes showed that in the first trials the number of changes was low, increased steeply up to the sixth three-minute trial, then slowly decreased. Similar results may not be found with other work groups where the organization has attained stability and methods do not change a great deal. Many formal organizations have a fairly slow turnover of people, so that at any given time, most of the members will know each other fairly well. Also, responsibility for developing new methods of performing a task usually is not given to the same people who perform the task. Finally, the task used in this study required close cooperation among all members in order for them to produce anything, whereas in many organizations which perform tasks, the group productivity may be merely the sum of the production of individuals.

---

[1] Unpublished manuscript.

SUMMARY

The relationship of the distribution of managerial traits within a group to productivity in a complex cooperative task was investigated, using groups of two, three, and four people. Two managerial traits, supervisory ability and decision-making approach, were measured by means of self-descriptions obtained from a forced-choice adjective checklist. Various combinations of these scores were formed for each group, and these were correlated with criteria of productivity and carefulness, the criteria being weighted composites derived through cluster analysis of various performance measures. It was found that neither the average amount of the trait possessed by the group nor the amount possessed by the highest scorer was highly related to the criteria. Two measures of skewness used, namely, the difference between the highest and next highest scorer, and this difference minus the range of the lower scorers, were highly related to the criterion of productivity, a correlation of .82 having been obtained between the criterion and the positive skewness of the group on the decision-making approach (DMA) scale. It is suggested that part of the relationship may be based on the fact that DMA positive skewness is negatively related to the number of organizational and methods changes occurring in the group, and the number of such changes is in turn negatively related to the productivity criterion.

REFERENCES

1. GHISELLI, E. E. The forced-choice technique in self description. *Personn. Psychol.*, 1954, **7**, 201–208.
2. PORTER, L. W., & GHISELLI, E. E. The self-perceptions of top and middle management personnel. *Personn. Psychol.*, in press.
3. TRYON, R. C. Identification of social areas from cluster analysis. *Univer. of Calif. Publ. Psychol.*, 1955, **8**, Nr. 1, 1–100.

Received June 18, 1957.

THEMATIC APPERCEPTION TEST: AN EMPIRICAL EXAMINATION OF SOME INDICES OF HOMOSEXUALITY[1]

GARDNER LINDZEY[2]
Syracuse University

CHARLOTTE TEJESSY
Massachusetts General Hospital

AND HAROLD S. ZAMANSKY
Northeastern University

MOST attempts to introduce systematic knowledge and precision of statement into the use of projective techniques are frustrated by an inability to identify a firm basis from which to initiate investigation. Each conceivable empirical problem immediately calls to mind a host of other problems that seem at least as significant and of prior importance. Even attempts to explore very elementary questions characteristically raise the shadows of dozens of other issues, the resolution of which seems necessary before the first question can be approached meaningfully.

Under such circumstances, one defensible tactic is simply to begin with a careful attempt to state explicitly the present status of projective testing or a particular projective technique, and then to examine piecemeal the accuracy or utility of the various elements in this position. This paper represents one part of such an approach to the study of the Thematic Apperception Test. We began with a statement of assumptions that underlie the conventional use of the test (5) and followed this with the identification of well over 500 generalizations that related specific aspects of Thematic Apperception Test response to some characteristic or attribute of the story teller (8). These statements were collated from the published literature and represent the efforts of clinicians and investigators to summarize and generalize their experience with the test. Given this list, we then set out to examine the empirical utility of significant clusters of the generalizations (6, 7). The present paper reports the results of several studies that examined the potential utility of existing indices of homosexuality.

We found initially a total of nine different generalizations concerning the relation between homosexuality and aspects of TAT response. Having identified these, we proceeded to develop a scoring procedure for each of them that would permit us to analyze TAT protocols objectively. We then related the resultant scores for a group of intensively studied subjects to various independent measures of sexuality. Further, we applied our derived scoring procedure together with some additional variables to the TAT protocols of a group of known, overt homosexuals as well as to the protocols of a normal control group.

RELATION BETWEEN TAT INDICES OF HOMOSEXUALITY AND RATINGS OF SEXUALITY

Our initial inquiry into the utility of the indices of homosexuality examined the association between these variables and several sets of ratings that were designed to assess the sexual motives of a relatively normal group of subjects.

Procedure

Subjects. The *S*s were 20 undergraduates of Harvard College who had volunteered from a large introductory psychology course to participate in an intensive program of personality study and who were paid at the customary rates for student employment. They were selected so as to be representative of their college class in terms of such variables as religion, socioeconomic status, academic standing, and extracurricular activities.

---

[1] This research was supported in part by Research Grant M-1949 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service. A portion of the research was carried out in connection with a program of investigation under the direction of Henry Murray, Harvard University, and supported by funds from Rockefeller Foundation. We wish to gratefully acknowledge the assistance of Dagny Kalnins and Philip Bossart in conducting these studies.

[2] Now at the University of Minnesota.

Ratings. The *S*s were studied over a two-year period by a team of roughly a dozen psychologists, using a wide variety of techniques that included virtually all of the conventional devices for assessing personality in addition to a number of newly devised techniques. Among these instruments were projective techniques, inventories, observational methods, autobiographical report, interviews, and various special ratings. At the end of the period of study, each *S* was discussed at length in a diagnostic council involving the various specialists participating in the study. Upon conclusion of this session, a general formulation of the individual's personality was agreed upon, and a number of ratings on specified variables were assigned. One of these ratings focused upon the variable of homosexuality. The judgment was made in terms of a six-point scale with the intention of reflecting the importance of this motive as a determinant of the *S*'s behavior. Inasmuch as a great deal of the information that the diagnostic council had at its disposal was presumed to bear upon covert or latent aspects of personality, it was natural that the ratings emphasized latent homosexuality. It was generally understood for all of the variables rated that if an individual showed strong and unmistakable evidence of the unconscious or covert operation of the motive, he would receive a high rating on the variable even though this motive did not receive direct expression in overt behavior. These ratings are referred to as *diagnostic council ratings*.

The fact that the diagnostic council judges were familiar with the TAT responses of the *S*s represents a possible source of contamination. The same factors that lead to high scores on the TAT indices of homosexuality might be attended to by the judges and thus lead them to equally high diagnostic council ratings on this variable. Actually, the judges did not have the results of any of these ratings nor were they specifically acquainted with the generalizations. Further, the amount of material that they had to consider and digest in arriving at their decisions made it unlikely that there would be any serious bias introduced by this factor. Finally, in two other studies (6, 7) it was demonstrated quite conclusively that the specific TAT indices were not mirrored in or important influences on the ratings of the diagnostic council.

In addition to the ratings of homosexuality, we also had available for 15 of the *S*s a set of *self ratings* concerned with heterosexual activity and needs. These ratings, also made on a six-point scale, reflected the *S*'s judgment of how strong was his need for heterosexual expression in comparison to other individuals of his age and general background. Finally, for 18 of the *S*s, we had a set of *observer ratings* of the same variable, made by an experienced clinical psychologist on the basis of a detailed autobiography and a fact-finding interview.

Thematic Apperception Test. The TAT was administered individually under standard conditions with the examiner taking notes and at the same time obtaining an electrical transcription of the stories. The two sources of information provided a virtually verbatim record of the *S*'s speech.

The stories were analyzed in terms of nine variables that had been derived from the collected list of empirical generalizations. In each case, we began with the published statement of the variable and then attempted to objectify the statement through the use of specific rules that would permit a quantitative outcome. In describing these nine variables we shall present a brief quotation that identifies the source of the variable, as well as its general content, and indicate the kind of rating scale that was used in analyzing the stories.

Discomfort on Card 9BM. "Card 9 is always significant because the picture of the men lying on the ground will make a male patient uncomfortable if he feels any attraction for other men" (1, p. 106). The scoring was dichotomous, with discomfort inferred from such indices as incidental remarks by the *S* indicating disturbance or an inability to think of a story, hesitation, and long pauses before beginning the story, extreme shortness of the story in relation to the typical length of the *S*'s stories.

Hypnotism on Card 12BM. "Stories to #12 in which the young man on the couch has given himself up to be hypnotized by the older man or in which the older man has forcibly hypnotized the young man . . . frequently reflect the patient's latent homosexual tendencies or overt homosexual experience" (11, p 46). Dichotomous scoring.

Misrecognition of Sex. "Latent homosexuality or a condition where a patient is in conflict concerning homosexual urges may express itself in distortions of the sex of the character in the pictures" (10, p. 90). Dichotomous scoring.

Attack From the Rear on Card 18BM. "Stories to #18 in which the hero has been attacked from the rear or pulled to the rear, frequently reflect the patient's latent homosexual tendencies or overt homosexual experience" (11, p. 46). Dichotomous scoring.

Feminine Identification. "Latent homosexuality or a condition where the patient is in conflict concerning homosexual urges may express itself in identification with female characters" (10, p. 90). Dichotomous scoring.

Attitude Toward Marriage. "Latent homosexuality or a condition where a patient is in conflict concerning homosexual urges . . . may also be discovered by basic attitudes toward marriage and the opposite sex which will occur with some frequency" (10, p. 90). Rated on a five-point scale.

Man Killing Woman. "Murray stated that men who produce stories in which men kill women are likely to have conflict over strong homosexual tendencies" (9, p. 114). Scored as present or absent with a borderline category for cases where a male has indirect responsibility for killing a woman.

Sexual References. "In manifest homosexuals I have noticed a tendency toward frequent manifest sexual references" (2, p. 216). Rated on a four-point scale.

Unstable Identification. "In manifest homosexuals I have noticed a tendency toward a recurrent shift of identification not only from males to females but also within one sex" (2, p. 216). Rated on a three-point scale.

For those variables that showed sufficient range, product-moment correlations were computed to estimate the degree of association between the experimental ratings and a set of independently performed reliability ratings. Three of these variables showed relatively high *reliability* (Attitude Toward Marriage, .83; Men Killing Women, .89; Sexual References, .88) while two showed only moderate reliability (Feminine Identification, .68; Unstable Identification, .63). The remaining four variables did not show sufficient range of scores to permit the usual correlation approach, but the use of a chi-square measure of association showed a highly significant covariance between the two sets of ratings.

Analysis of data. In relating the TAT indices to the various ratings, the product-moment correlation was employed wherever the distribution of the data permitted; in all other instances, the biserial correlation was used. In each case, the direction of the relationship was predicted in advance, and for this reason a one-tailed test of significance was used consistently.

TABLE 1

RELATION BETWEEN RATINGS AND TAT INDICES OF HOMOSEXUALITY

TAT Index	Ratings					
	Diagnostic Council $n = 20$		Observer $n = 18$		Self $n = 15$	
	r	r_{bis}	r	r_{bis}	r	r_{bis}
1. Discomfort on Card 9BM		−.30		.16		−.31
2. Hypnotism on Card 12BM		−.33		.32		−.04
3. Misrecognition of Sex		.12		−.05		−.33
4. Attack From the Rear on Card 18BM		−.11		.00		−.31
5. Feminine Identification		.48*		−.11		−.24
6. Attitude Toward Marriage	.01		.47*		.41	
7. Man Killing Woman	.02		.25		−.37	
8. Sexual References	.49*		.18		.13	
9. Unstable Identification		.22		.09		−.22

* Significant at the .05 level of significance.

Results

It was our general expectation that the TAT signs would show the highest degree of association with the diagnostic council ratings of latent homosexuality and the least association with the self ratings of heterosexuality. This prediction was based on the assumption that the outcome of the complex diagnostic council procedure was a more accurate measure of the *S*'s latent motivation than the judgments of the single observer, which in turn should be superior to the self ratings. We expected, of course, that the ratings of latent homosexuality would be positively related to our TAT indices, while the ratings of heterosexuality would be negatively related to these indices. It is apparent that the above reasoning is grossly oversimplified, as there are many complicating factors that would be expected to operate in this situation which we have not considered. Nevertheless, in advance of the study this seemed the most defensible general prediction that could be made.

The results of our analysis are summarized in Table 1, where we find a weak confirmation of our expectation that the TAT indices would show evidence of positive association with the diagnostic council ratings of homosexuality. Of the nine correlations, six are positive as predicted (although two of these are of negligible magnitude), and two attain significance at the .05 level. None of the three negative correlations attain conventional significance. The observer ratings clearly provide an even less impressive correlational picture. Only two of the nine correlations are in the predicted direction, and the single coefficient that attains significance is not one of these. Therefore, on grounds of number of significant correlations, as well as on grounds of the general tendency of the data to fall in predicted directions, the observer ratings seem less intimately related to the TAT indices than are the diagnostic council ratings, as we had predicted.

There are no significant correlations between the self ratings and the various TAT indices, but there is more consistency in direction here than in either of the other cases. Seven of the nine correlations are in the predicted direction, and four of these are above .30 as compared to only two of the correctly predicted relations involving the council ratings. Thus, our hypothesis that the self ratings would show the least evidence of association with the TAT indices does not appear to be confirmed.

Although these relations are sufficiently slight so that it is difficult to assert anything with conviction, it seems clear that the variables concerned with Feminine Identification and Sexual References merit further investigation. Worth noting is the fact that three of the indices of homosexuality most frequently encountered among users of the TAT (Discomfort on Card 9BM, Hypnotism on Card 12BM, Attack From the Rear on Card 18BM) actually show a negative association with the council ratings. There is no evidence here to support custom! This observation, when coupled with the finding that the same three variables were related to the self ratings in the predicted direction, is reminiscent of the results of a similar study (7) where it was discovered that TAT indices of aggression were most intimately related to self ratings and least closely associated with diagnostic council ratings.

In this earlier study it was found that a general clinical rating, based on the TAT protocols with no other knowledge of the *S* and with no attempt at objectivity and careful quantification, led to a correlation with the council ratings that was as high as the best of the TAT indices. Virtually the same finding was unearthed in the present study. When general, clinical ratings of homosexuality, made by the individual who administered the test and with no further information concerning the *S*, were related to the council ratings, they showed a product-moment correlation of .47. This is approximately the same as the highest of the correlations obtained with the carefully scored TAT indices. Thus, the painstaking efforts at quantification seem to have produced an end result which, on the average, is inferior to the easy, general rating.

Before considering any further implications of these findings, let us turn to a second study that employed many of these same TAT indices and in this case scrutinized them against a somewhat more sharply differentiated criterion than that used in the present study.

DIFFERENCES BETWEEN OVERT HOMOSEXUALS AND NORMAL SUBJECTS ON TAT INDICES OF HOMOSEXUALITY

Having examined the relation between our TAT indices and various ratings of sexuality for a normal group of *S*s, the question next arose as to how well these indices would function when examined in connection with a broader range of sexual behavior. In order to answer this question, we obtained a set of TAT protocols from a group of homosexual *S*s and a control set from a group of normal *S*s and compared the performance of the two groups on all of the variables used in the previous study that could be applied to the present protocols. We also added to our comparison a number of variables that we had encountered in the literature subsequent to the first study, and we devised certain new variables in the expectation that they might differentiate the two groups. Finally, we classified the two groups of *S*s on the basis of a clinical appraisal of the TAT stories, making no effort to provide ratings or specific justification for these decisions.

Procedure

Subjects. The *S*s consisted of 20 undergraduate males who had admitted to overt homosexual acts and 20 undergraduate males who to the best of our knowledge had never committed a homosexual act. None of the *S*s displayed any gross evidence of psychopathology, and most of the homosexual group had accepted their sexual deviation and were not receiving psychotherapy. The two groups were matched in sex, age, and educational level. The reader should note that a small number of our homosexual *S*s were also included in a study reported by Davids, McArthur, and Joelson (3); and consequently, the findings of these two studies can not be considered cumulative or to be independent estimates of the actual relationships existing.

Thematic Apperception Test. A shortened form of the TAT, consisting of Cards 4, 6BM, 7BM, 10, and 18BM, was administered individually by a male administrator. The protocols were typed uniformly so that the rater was unaware of the identity of any of the *S*s when analyzing the stories. The protocols were then scored for the following variables taken from the list used in the previous study: Misrecognition of Sex, Attack From the Rear on Card 18BM, Feminine Identification, Attitude Toward Marriage, Man Killing Woman, Sexual References, and Unstable Identification. In a few cases, the scoring system was slightly modified, but in general these stories were scored in very much the same manner as the first set. The first two variables of the previous study (Discomfort on Card 9BM and Hypnotism on Card 12BM) could not be used as they are linked to specific TAT cards not employed in this study. However, a number of additional variables were examined, and these are briefly described below together with an indication of the source of the variable and the general form of the rating.

Representing Feminine Feelings and Emotions. "... evidence of his having experienced many things in a 'feminine' way himself ... ability to create a picture of (a woman's) yearnings and feelings (card 10) ..." (4, p. 215). Rated on a five-point scale.

Shallow Heterosexual Relations. "... patients who are under the pressure of strong homosexual needs, usually paranoids, and neurotic characters who cannot conceive of a genuine tender relationship between any two persons ... think that the scene (Picture 10) is a sham, that there is really no reciprocal affection" (4, p. 216). Scored on a five-point scale ranging from positive and satisfying relation to relation lacking in love or affection.

Male Embrace on Card 10. "... a strong clue to latent homosexuality or even manifest problems of this nature.... If this (Picture #10) is interpreted as an embrace between males by a male subject" (2, p. 210). Scored dichotomously.

Attitude Toward Opposite Sex. In the first study, attitudes toward marriage and the opposite sex were scored as a single variable, whereas in the present study these were separated and scored individually on a five-point scale ranging from very positive to very negative.

Tragic Heterosexual Relations. "The homosexuals gave several stories in which a *heterosexual relationship was followed by tragedy*, usually the violent and sudden death of one member of the couple" (3, p. 168). Scored dichotomously.

Attachment to Father. "The first sign (of homosexuality) ... applies to stories containing a *strong unresolved attachment to a father or father figure* composed for card 8" (3, p. 168). There was a typographical error in the original article, and this sign should be referred to Card 7BM. In the present study, this variable was scored dichotomously for Card 7BM and Card 18BM.

Derogatory Sexual Terms Applied to Women. "Although the homosexuals did not often produce stories in which a clearly derogatory attitude toward women was revealed, use of certain *sexual derogatory terms in relation to women* gave specific clues to their feelings toward female characters in their stories" (3, p. 168). Scored dichotomously.

Attachment to Mother. "Another sign (of homosexuality), associated with Card 6, applied to stories involving a *strong unresolved attachment to the mother*" (3, p. 168). This variable was scored dichotomously for both Cards 6BM and 10BM.

Symbolism or Allegory in Response to Card 18BM. "... in response to Card 18, several homosexuals told stories that *describe the plight of the young man allegorically or symbolically* ..." (3, p. 168). Scored dichotomously.

In addition to these variables, certain further indices were formulated in the expectation that they might reveal differences between the two groups.

Incest. This variable was scored as present or absent, with presence being assigned to any kind of sexual relation among members of the same family.

Manifest Homosexuality. Scored dichotomously with homosexual act, thought, impulse, or even use of the word being scored as presence of the variable.

Perception of Card 10 as Other Than Elderly Couple. Scored dichotomously.

Introduction of Female in Positive Context on Card 18BM. In this case, it was expected that the normal group would show the higher incidence of the sign which was scored as present or absent.

Reliability. In order to assess the extent to which these indices could be scored reliably, a second rater with no shared practice with the experimental rater performed the same ratings on 40 stories selected so that there was one story from each of the *S*s and an equal number of stories told to each of the pictures employed in this study. The relation between the scores of the two raters is summarized in Table 2. Where the data permitted, product-moment correlations were computed, and where the range of scores was insufficient to permit this, tetrachoric correlations were computed. In those instances where there were too few entries in one or more of the cells to permit a reasonable use of the tetrachoric correlation technique, we have simply reported the percent of complete agreement between the two raters. In general, the results are highly satisfactory. The lowest Pearson r is .71, the lowest tetrachoric correlation coefficient is .81, and the lowest percentage of agreement is

TABLE 2

INTERRATER RELIABILITY OF TAT INDICES

($n = 40$)

Variable	% Complete Agreement	r	r_{tet}
Misrecognition of Sex	91		
18BM: Attack From the Rear			.84
Feminine Identification	83		
Attitude Toward Marriage		.71	
Man Killing Woman	98		
Sexual References	98		
Unstable Identification			.81
Feminine Feelings, Emotions		.83	
Shallow Heterosexual Relations		.89	
Male Embrace	100		
Attitude Toward Opposite Sex		.84	
Tragic Heterosexual Relations	100		
Attachment to Mother			.84
18BM: Symbolism or Allegory	92		
Attachment to Father	92		
Derogatory Sexual Terms Applied to Women	97		
Homosexual Content	97		
Incest	100		
10BM: No Elderly Couple	88		
18BM: Positive Introduction of Female	95		

83. Considering the fact that this degree of agreement was achieved without any shared experience on the part of the two raters, the results compare very favorably with those reported by other investigators.

Analysis of data. In the analysis of these data, the differences between the two groups were tested through the use of conventional t tests where the data permitted. In those cases where the data departed so far from normality as to prohibit this, we simply counted the occurrence of the index among the *S*s and tested the significance of the difference between the resulting proportions of occurrence of the index in the two groups. In a few cases, the theoretical frequency in one or more of the cells was less than five, and in these cases Fisher's Exact Test was employed. A one-tailed test of significance was used consistently because we predicted the direction of the relationship in every case.

Results

The differences between the homosexual and normal groups for the 20 variables are summarized in Table 3. It is clear that the data generally tend to bear out our prior expectation as only 3 of the 20 differences reverse the predicted direction. Moreover, a considerable number (nine) of the differences that conform to our prediction attain significance at the .05 level of significance or better, while none of the relations reversing our prediction attained conventional significance.

The nine variables that successfully differentiated between the criterion groups were Shallow Heterosexual Relations, Feminine Identification, Unstable Identification, Attitude Toward Marriage, Man Killing Woman, Use of Symbolism or Allegory on Card 18BM, Derogatory Sexual Terms Applied to Women, Homosexual Content, and Introduction of a Woman in a Positive Context on Card 18BM. In general, these variables yield a consistent picture of negative or hostile sentiments toward members of the opposite sex and psychological femininity.

It is immediately clear that our variables are more intimately related to the distinction between overt homosexuals and normal *S*s than they are to the various independent ratings of sexuality. Such a finding could reflect either the infirmity of our ratings, the greater range of sexual behavior involved in the second study, or the tendency of these signs to discriminate most effectively overt manifestations of the motive. In a study of indices of aggression (7), there were quite clear-cut findings indicating that the TAT indices were more closely related to conscious than to latent or unconscious aspects. That the case is not quite so simple in the present inquiry is indicated by the failure in the rating study to find any clear evidence of a closer relation between the TAT indices and the self ratings than between the TAT indices and the diagnostic council ratings.

When the two variables that were significantly related to the diagnostic council ratings in the first study are examined in the present context, we find that both show some evidence of differentiating the groups in the predicted direction. This trend is very slight and nonsignificant in the case of Sexual References but is highly significant for the variable of Feminine Identification.

What of the remaining variables that were employed in both studies? The evidence is completely consistent in regard to the lack of utility of Misrecognition of Sex. The studies agree in suggesting the relative promise of Unstable Identification and Man Killing Woman. The findings are relatively indeterminate in both cases for the variable of

TABLE 3

DIMENSIONAL TAT COMPARISON OF NORMAL AND HOMOSEXUAL SUBJECTS

Variable	Normal		Homosexual		$s_{\bar{x}}$	t	χ^2	p
	$\bar{X}$	Freq.	$\bar{X}$	Freq.				
Misrecognition of Sex		3		3				—
18BM: Attack From the Rear		4		9			1.82	$<.10$
Feminine Identification	1.05		1.55		.11	4.46		$<.005$
Attitude Toward Marriage	−.32		.48		.28	2.82		$<.005$
Man Killing Woman		0		5				$<.02$
Sexual References	.55		1.25		.22	.32		$<.40$
Unstable Identification	1.65		2.25		.13	4.76		$<.005$
Feminine Feelings, Emotions*	2.10		2.05		.30	.17		$>.50$
Shallow Heterosexual Relations	.90		1.75		.13	6.24		$<.005$
Male Embrace		0		1				—
Attitude Toward Opposite Sex	−.32		.18		.77	.65		$<.30$
Tragic Heterosexual Relations		2		7				.06**
Attachment to Mother		2		4				.33
18BM: Symbolism or Allegory		0		6				.01
Attachment to Father*		5		1				.18
Derogatory Sexual Terms Applied to Women		0		7				.004
Homosexual Content		0		6				.01
Incest		0		1				—
10BM: No Elderly Couple*		16		11			.46	$<.50$
18BM: Positive Introduction of Female		5		0				.02

* These are the only variables where the difference was *not* in the predicted direction and consequently the only variables where a two-tailed test of significance was employed.

** Where there is no value for either t or χ^2 the probability estimate was derived from Fisher's Exact Test.

Attack From the Rear, while the variable of Attitude Toward Marriage functioned well in the present study and very poorly in the rating study. In summary, the two studies produced relatively congruent findings for six of the seven overlapping variables.

These indices of homosexuality have functioned relatively well against the criterion measure. Certainly this is the case when they are compared to the effectiveness of indices of anxiety and aggression. One may still question seriously, however, whether any or all represent much of an aid for the clinician or investigator. It is evident that the variables that function best are precisely those which one would not expect to encounter when dealing with strongly repressed homosexuality or with an overt homosexual who was attempting to conceal this aspect of his behavior. Indices having to do with overt homosexuality, crude sexual references, derogatory remarks concerning the other sex, etc., are all readily susceptible to conscious censoring. While it remains an interesting research question as to what the differences are between TAT responses collected under circumstances where the examiner is known to be familiar with the homosexuality of the *S* and under circumstances where the examiner is not aware of this, there seems little doubt that many of the present characteristics of the homosexual protocols would be minimized or eliminated where the deviance of the *S*s was less openly maintained.

A different approach to the clinical utility of these indices can be made through comparing the results of our objective analysis with the results of a general clinical sorting of the 40 protocols. This categorizing was made by one of the authors without awareness of the identity of any of the *S*s but with a general familiarity with all the specific variables used in the study. He was able to sort the protocols with 95 per cent accuracy. In other words, he made only one reversal, classifying a single homosexual protocol as normal and a single normal protocol as homosexual. The accuracy was actually even greater than this over-all figure implies, because in making these ratings, he initially sorted the stories according to his certainty or confidence of their classification (16 in the confident homosexual group and 13 in the confident normal group), and all of these were identified correctly. Further, the stories were ranked in terms of the certainty of the judgment so that a rank of 20 would represent the set of stories that was least confidently classified. The two stories misclassified were respectively 19th in the homosexual group and

16th in the normal group. In other words, the stories were correctly classified in 95 per cent of the cases, and the two errors were made on judgments held in relatively low confidence by the judge.

Clearly, the uninhibited clinician functions better than any one of the objective indices, and furthermore he functions better than any reasonable combination of these indices. This observation is not meant to disparage the indices but simply to point out again the relative inefficiency of customary indices or signs as diagnostic tools.

DISCUSSION

The results of our two inquiries agree in suggesting that those aspects of the TAT that have been identified conventionally as resonant to homosexuality are approximately as advertised. That is, there is evidence of association in the predicted direction between our TAT indices and the criterion measures. The strength or magnitude of the observed relations is generally not very high, but there can be little question that these indices have functioned better than equivalent sets of indices for the variables of anxiety (6), aggression (7), and schizophrenia examined in past studies.

Our tentative findings suggest that the TAT protocols of the individual characterized by strong homosexual tendencies display consistently negative attitudes toward women, a lack of full, rich, and satisfying relations with members of the opposite sex, and occasional reference to manifest homosexuality. In addition, there were a few formal differences between the stories of the normal and the homosexual, such as the tendency on the part of homosexual *S*s to utilize symbolism or allegory on Card 18BM. As we have already indicated, there is no certainty that these signs will continue to function effectively in settings that differ from those employed in the present study. Not only does this caution apply to differences involving *S* awareness of the examiner's knowledge of his homosexuality but, also, to variation in such characteristics as intelligence, educational level, and group membership. Thus, the use of allegory or symbolism seems almost certain to be a function of homosexuality only under circumstances where a certain minimum of intelligence and education has been attained.

What of the distinction between latent and overt homosexual tendencies? Here our evidence is rather inconclusive but generally points to the rather discouraging conclusion that the TAT was most diagnostic of homosexual tendencies by means of relatively manifest variables and in relation to an overt homosexual criterion. That is, most of the indices of homosexuality that functioned successfully tended to be rather directly related to homosexuality and thus might be expected to be readily subject to censoring or inhibition. Further, the signs showed a much more convincing association with the criterion provided by the homosexual and normal groups than by the diagnostic council ratings of homosexuality. This advantage of the overt homosexual criterion could, of course, reflect the relative purity or sensitivity of the two criterion measures rather than the difference between them along the covert-manifest continuum.

Perhaps the strongest congruence provided by our two studies bears upon the comparison of general or clinical judgments and the objective TAT indices. In both investigations it was evident that the uninstructed and unhampered clinician could do a much better job of predicting the criterion than could the diligent rater. When this consistent finding is added to a similar observation made in a study of indices of aggression (7), the implication seems undeniable that our difficult and time consuming ratings are not functioning at a high level of efficiency. It is not surprising that early efforts at objectivity and specification should lead to some diminution in clinical effectiveness, but this finding does make clear the limited utility of the TAT indices at present.

SUMMARY

We have conducted two studies designed to estimate the effectiveness of a variety of TAT indices that have been proposed as measures of homosexual tendency. In the first study, nine TAT indices were scored for 20 subjects and related to independent ratings derived from a diagnostic council, an objective observer and the subject himself. Results indicated that the indices appeared more closely related to the council ratings than to observer ratings, but there was no evidence of any superiority on the part of the observer ratings over the self ratings. A general clinical rating based on the TAT proved to be more closely related to

the diagnostic council rating than any of the objective indices.

In the second study, 20 TAT indices of homosexuality were examined for their effectiveness in differentiating between 20 overt homosexual subjects and 20 normals. The findings revealed 16 of the signs differentiating between the groups in the predicted direction, while three reversed the prediction and one showed no difference. Nine of the 16 predicted differences were significant at the .05 level, whereas none of the reversals attained conventional significance. A judge unacquainted with the identity of any subject was able to sort the protocols into homosexual and normal groups with 95 per cent success. This represents a much more efficient discrimination than that permitted by any of the objective indices.

Although these indices of homosexuality have functioned more effectively than equivalent indices for other variables, there still seems ground for serious doubt concerning their utility. Not only does the general clinical rating appear to function more effectively, but also the nature of the indices implies that they could easily be subjected to voluntary distortion or inhibition, thus minimizing their usefulness in many diagnostic settings.

REFERENCES

1. ARNOLD, MAGDA B. A demonstration analysis of the TAT in a clinical setting. *J. abnorm. soc. Psychol.*, 1949, **44,** 97–111.
2. BELLAK, L. The Thematic Apperception Test in clinical use. In L. E. Abt & L. Bellak (Eds.), *Projective psychology: clinical approaches to the total personality*. New York: Knopf, 1950. Pp. 185–229.
3. DAVIDS, A., JOELSON, M., & MCARTHUR, C. Rorschach and TAT indices of homosexuality in overt homosexuals, neurotics, and normal males. *J. abnorm. soc. Psychol.*, 1956, **53,** 161–172.
4. HOLT, R. R. The Thematic Apperception Test. In H. H. Anderson & Gladys L. Anderson (Eds.), *An introduction to projective techniques*. New York: Prentice Hall, 1951. Pp. 181–229.
5. LINDZEY, G. Thematic Apperception Test: Interpretive assumptions and related empirical evidence. *Psychol. Bull.*, 1952, **49,** 1–25.
6. LINDZEY, G., & NEWBURG, A. S. Thematic Apperception Test: A tentative appraisal of some "signs" of anxiety. *J. consult. Psychol.*, 1954, **18,** 389–395.
7. LINDZEY, G., & TEJESSY, CHARLOTTE. Thematic Apperception Test: Indices of aggression in relation to measures of overt and covert behavior. *Amer. J. Orthopsychiat.*, 1956, **26,** 567–576.
8. LINDZEY, G., TEJESSY, CHARLOTTE, DAVIDS, A., & HEINEMANN, SHIRLEY H. Thematic Apperception Test: A summary of empirical generalizations. Unpublished manuscript, Harvard Univer., 1953.
9. PIOTROWSKI, Z. A. A new evaluation of the Thematic Apperception Test. *Psychoanalyt. Rev.*, 1950, **37,** 101–127.
10. ROTTER, J. B. Thematic Apperception Test: suggestions for administration and interpretation. *J. Pers.*, 1946, **15,** 70–92.
11. STEIN, M. I. *The Thematic Apperception Test: An introductory manual for its clinical use with adult males.* Cambridge: Addison-Wesley, 1948.

Received June 19, 1957.

THEMATIC APPERCEPTION TEST: SOME EVIDENCE BEARING ON THE "HERO ASSUMPTION"[1]

GARDNER LINDZEY[2] AND DAGNY KALNINS
Syracuse University

FROM its very moment of origin, the Thematic Apperception Test has been intimately associated in the minds of most users with the assumption that there are certain characters in each story that clearly reflect attributes of the storyteller, while other figures are more revealing of the storyteller's perceptions of the individuals who populate his personal world. In an empirical discipline, however, the length of time that a belief has been held entitles it to no special consideration, and it is therefore quite natural that this assumption should be challenged.

Objections to the hero assumption were particularly prone to occur because of the intimate relation between an individual's own attributes and that which he perceives in the outer world. Thus, it is commonly assumed, with considerable supporting evidence, that an individual who is highly aggressive will perceive more hostility in the world around him than an individual who is less aggressive. This association between internal and external tends to disrupt and blur the easy distinction implied by the original assumption. It should be mentioned that Murray, who initially formulated this assumption, was fully aware of the complexity of the relation between the individual's inner and outer worlds and never intended that the test interpreter should slavishly maintain a rigid distinction between "hero" and "other." In fact, the manual originally published with the test (12) contains a lengthy discussion of various complications that under special circumstances make it necessary to modify or abandon the assumption of a single hero or identification figure.

Even accepting the fact that ultimate precision will surely depend upon some more complicated set of assumptions than those we are here concerned with, it is still an interesting question whether the interpreter of the TAT can go further with the assumption of a distinction between hero and non-hero than he can with the assumption that all figures in the story are equally revealing of the "own characteristics" of the storyteller. Although endless rational arguments can be introduced bearing on the choice between these assumptions, what we need most is not logical inference nor emotional polemic but controlled empirical evidence. In this spirit, the present paper is intended to outline the results of two small investigations designed to provide some very tentative findings that bear upon the utility of employing the distinction between hero figures and non-hero figures in interpreting Thematic Apperception Test protocols.

In the first study, we began with the very simple notion that if "heroes" were more indicative of characteristics of the subject (S) than were "other figures," the S should perceive them or react to them differentially. Given this reasoning, we proceeded to administer the TAT to a number of Ss, conducted an individual inquiry in which we tried to assess the reaction of the storyteller to each of the figures in his stories, and then looked for differences in reaction to those figures independently rated as identification figures as opposed to those that were not rated as identification figures.

In the second study, we were able to examine certain quantitative shifts that occurred in TAT stories following a frustration experience within the framework of the hero assumption and within the framework of the assumption that all figures are equally indicative of characteristics of the storyteller.

It should be clearly understood that in these studies we are *not* asking whether the simple assumption of a single hero in each story is the most fruitful assumption that can be made. What we *are* asking is whether this is a more fruitful assumption than the opposite extreme, the assumption that all figures in the story are

---

[1] This investigation was supported by Research Grant M-1949 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service. We are grateful for the assistance of Willa D. Abelson in the execution of this research.

[2] Now at the University of Minnesota.

equally revealing of the storyteller's characteristics. There are many fine gradations between these extremes, and a shrewd observer with a facile pen can complicate the assumptions endlessly. Some of the many alternative assumptions have been considered in an earlier paper by one of the present authors (6), and his presentation has been examined critically by Piotrowski in several subsequent papers (13, 14). What is needed now, however, is not further complexity or rational elaboration but rather a statement of assumptions with sufficient explicitness so that they lead to clear empirical consequences, followed by a careful testing of these consequences.

The findings presented here are, at most, a beginning on the road to clarifying the kinds of underlying processes that operate in the construction of imaginative stories. Their sole virtue lies in the fact that they make clear under controlled circumstances something about the relative merit of two widely divergent and yet defensible assumptions concerning the process of interpreting imaginative protocols.

SUBJECT REACTIONS TO HERO AND NON-HERO FIGURES

If the assumption of a hero in each story whose attributes are especially revealing of the *S*'s psychological makeup is to prove defensible, it seemed to us that we should be able to show that the *S* reacted differently to hero figures than he did to those figures judged not to be hero figures. In particular, we reasoned that if heroes were carriers of storyteller attributes, the *S* should see these figures as more similar to himself than the non-hero figures, or else he should react with a violent denial of any similarity or resemblance between himself and the hero figures. This conclusion was derived from the assumption that the TAT revealed both conscious attributes and unconscious or unacceptable attributes of the *S*, coupled with the further reasoning that when a figure represented a conscious quality of the storyteller, he would accept and report the similarity, while under circumstances where there was an unacceptable impulse or quality involved, he would tend to deny strongly any similarity between himself and the figure.

Consequently, the hypothesis to be tested in this study asserts that story-characters independently judged to be hero figures are seen by the *S* as similar to himself or as having no self-similarity whatsoever, while figures independently judged *not* to be hero figures are more often seen by the storyteller as resembling others or else as representing stereotyped or fictional characters. It seems clear that the assumption that all figures are equally revealing of the storyteller provides no basis for predicting any difference in *S* reaction to hero and non-hero figures.

Procedure

A shortened version of the TAT consisting of Cards 2, 5, 7GF, 9GF, 10, and 18GF was administered in a small group setting to 30 Syracuse University undergraduate females. Use of group administration seemed warranted in view of the results of an earlier investigation (7) comparing individual and group administration. The *S*s had volunteered to participate in the study from an introductory course in psychology. After completion of the group test, individual appointments were made for each *S* within 48 hours from the time of the original test. During the individual interview, each *S* was asked to tell as much as she could concerning the factors that led to her creating each of the stories she had constructed. In addition, for each character in each of the stories, she was required to make a judgment of how similar the character was to herself and how similar it was to other people whom she had known. The responses of the *S*s permitted us to categorize each character in terms of similarity to self (thought of as self, could be self, some similarity to self, denial of similarity to self) and similarity to other (thought of as other, similar to other). The characters were also classified in a general stereotype category when the *S* reported that the figure represented some general fictional character or when she reported no resemblance to anyone and could not say what had influenced her to create this character.

The stories were independently analyzed in order to identify the hero and non-hero figures in each story. Two raters with no shared practice and with only a few general scoring principles were able to reach complete agreement on the identity of the hero figure (or the absence of a hero) in 90 per cent of the 180 stories rated.

The analysis of these data posed certain thorny problems as the most obvious and pertinent arrangements of the data led to more observations than subjects and thus made customary statistical analysis inappropriate. In view of this difficulty, we followed the convention of presenting the data descriptively in what seems to us the most revealing manner and then, in addition, performing certain further analyses that permit the application of the usual tests of significance.

In order to reduce the number of observations to the number of *S*s, we adopted a procedure for each of the three relevant areas of response (similarity to self rather than other, denial of similarity to self, stereotype) that permitted us to assign to each *S* a score representing the extent to which his six stories fitted with or deviated from the predicted pattern. Thus, if the first story that the individual told found him identifying the judged hero of the story as resembling himself, while the non-hero figures were judged to be similar to others, a score of one would be assigned. If a non-hero figure was judged to be similar to the self, while the hero figure was not judged to be similar to the self, a score of minus one was assigned. If the hero and non-hero figures were treated in identical fashion, a score of zero was entered. The total score for each *S* consisted of the sum of the individual scores for his six stories and could theoretically vary from plus six (confirmation of the hypothesis in every story) to minux six (negation of the hypothesis in every story). This procedure provided three sets of scores representing the extent to which the stories and inquiry responses of the individual *S*s conformed to our predictions in regard to similarity to self versus other, denial of similarity to self, and stereotyped response.

Results

The general results of the study are summarized in Table 1, where we find the distribution of hero and non-hero figures in each of the categories having to do with similarity to self and to other as well as the stereotype category. A comparison of the frequencies actually obtained with those that would be expected by chance alone makes it evident that the trend of these data is in support of the predictions made in advance of the study. Figures independently rated as heroes tend to be perceived as more similar to the storyteller than non-hero figures, or else are denied any similarity to the storyteller. On the other hand, non-hero figures, when compared to hero figures, tend more often to be identified as similar to some person other than the storyteller or else are classified as stereotyped. If the categories implying various degrees of similarity to self are combined into a single category and the same operation is performed for categories representing similarity to others, it is then possible to examine the association between the hero and non-hero distinction and similarity to self and other in a single 2 × 2 table. In Table 2, the outcome of such a procedure is represented, and there is clear evidence for association between the hero designation and perceived similarity to the self. It is, of course, not legitimate to perform the usual statistical analyses because of the lack of independence of observations. However, any such test applied to this array of findings would indicate the predicted association at a highly significant level. In general, these findings hold true not only for the over-all distribution reported in Table 1, but they are also sustained when the same distribution for each of the six cards is examined separately.

TABLE 1
SUBJECT REACTIONS TO HERO AND NON-HERO FIGURES

Reactions	Frequencies				Total
	Hero		Non-Hero		
	Observed	Chance Expectancy	Observed	Chance Expectancy	
Self	3	2.46	3	3.54	6
Could be self	14	7.38	4	10.62	18
Similar to self	21	11.07	6	15.93	27
Denial of self	29	24.19	30	34.81	59
Other	9	11.07	18	15.93	27
Could be other	33	36.08	55	51.92	88
Stereotype	78	96.76	158	139.24	236
Total	187		274		461

TABLE 2
SIMILARITY TO SELF AND OTHER OF HERO AND NON-HERO FIGURES

	Hero	Non-Hero
Similar to self	38	13
Similar to other	42	73

The results of our statistical comparison of hero and non-hero figures are presented in Table 3, where it is again made clear that the data tend to support our hypothesis. When heroes and non-heroes are compared in regard to *perceived similarity to self and others*, the scores present a significant deviation from chance in favor of the predicted greater similarity of self to hero figures and of others to figures judged to be non-heroes. The stories of 18 of our *S*s revealed the predicted association between self-similarity and hero figures and other-similarity and non-hero figures; for four of the *S*s, the hero and non-hero figures were treated in similar fashion, while the remaining eight *S*s reversed the prediction. Examination of the tendency to *deny similarity to self* reveals confirmation of our prediction that this would occur more frequently with hero than with non-hero figures. A number of *S*s provided no evidence of denial in any of their responses, so there was no possibility of differential perception of the hero and non-hero figures; but in those cases where there was such evidence, 12 of 17 *S*s showed a tendency to react with denial to hero figures more often than to non-hero figures. We also found evidence that non-hero figures were more often categorized

TABLE 3
CONFIRMATION OF PREDICTIONS CONCERNING DIFFERENCES IN REACTION TO HERO AND NON-HERO FIGURES

Reaction	$\bar{X}$	$s_{\bar{x}}$	t	p
Hero similarity to self and non-hero similarity to other	.966[a]	.343	2.82	<.05
Denial of similarity to self	.343	.186	2.31	<.05
Sterotype	1.033	.261	3.95	<.01

[a] A positive deviation from zero indicates confirmation of the predicted relation.

TABLE 4
REPORTED SOURCE OF TAT STORIES

Source	Frequency of occurrence	
	n	%
Imagination	87[a]	48
Properties of card	66	37
Autobiographical event	37	21
Experience of others	30	17
Reading: fiction	28	16
Movies, TV	24	13
Reading: non-fiction	8	4

[a] Total number of stories: 180.

as *stereotyped* or unrelated to either self or to other persons than were hero figures. There were nine *S*s who revealed no difference in the incidence of hero and non-hero figures who were perceived as stereotyped; but of the remainder, 17 reported the non-hero figures as more stereotyped, and only four saw the hero figures as more stereotyped.

The above findings provide clear confirmation of our predictions and thus support the value of the hero assumption. These same results, however, dramatically underline the shortcomings of this assumption in the face of certain TAT stories. Although the general trend of the data fits with the derivation from the hero assumption, there are individual *S*s who consistently reverse the prediction; e.g., there are some *S*s who characteristically view non-hero figures as like themselves and hero figures as like others. Thus, it seems clear that the actual situation is more complex than a literal application of the hero assumption would imply. Consequently, it becomes an important investigative task for the future to discover something about the types of *S*s or stories or both where this assumption may be applied fruitfully, as well as those where some other assumption should be utilized. This same conclusion is supported by the general findings contained in Table 1. While these results support our prediction, it is a group trend that we observe with many individual exceptions.

A very brief summary of our *S*s' reports concerning what had provided them with the idea for their stories is contained in Table 4. The results summarized here make clear that in a large number of cases (48 per cent) the *S*s indicate only that the story came from their imagination. Next most frequently mentioned as a determinant is the picture or some specific element within it (37 per cent). In those cases where the *S*s are able to identify a specific experience as having suggested the story, this was less likely to have been a fictional encounter (movies 13 per cent, novels 16 per cent) than an autobiographical event (21 per cent) or general experience and observation (17 per cent). Only a very small number (4 per cent) of the stories were reported to have stemmed from nonfiction reading.

In summary, the general findings of this study provide modest support for the hero assumption, although they also suggest that in individual cases the data do not mesh smoothly with this assumption. Let us turn now to the second study.

CHANGES FOLLOWING FRUSTRATION VIEWED FROM THE VANTAGE OF THE HERO ASSUMPTION

As we have already agreed, one of the major difficulties in evaluating the general utility of the hero assumption is posed by the close relation between internal states of the *S* and his perception of external reality. It is not easy to conceive of circumstances that on a priori grounds seem likely to produce changes in the motivational state of the *S* which will not be mirrored in changes in his perception of external reality. What we need is to construct a set of conditions under which the hero assumption predicts differential changes in hero and non-hero characteristics while the other assumption, of course, predicts consistent changes for both types of figures.

Such a combination of conditions seemed to exist in connection with a set of data that was part of an earlier study (5). These data consisted of a set of TAT protocols collected before and after a frustration experience together with an appropriate set of control protocols. Most important, a great deal was known about the details of the frustration experience and the *S*s' reactions to this situation. This information permitted us to make specific predictions concerning the changes that could be expected if one assumed all figures to be equally representative of the storyteller or if one assumed that there was only one figure, the hero, that reflected personal characteristics of the storyteller.

In brief, the frustration situation was of such a nature that it seemed plausible to expect that the *S*'s hostility toward others would increase, that he would also perceive others as directing hostility or aggression toward him, and finally that he would feel guilty or direct aggression toward himself. All acts of aggression within the TAT stories were analyzed in terms of whether the aggression was directed from (*a*) hero against other, (*b*) hero against self, (*c*) other against hero, or (*d*) other against other.

The assumption that all figures were equally characteristic of the storyteller implied that there should be a significant increase in aggressive acts of all four types in view of our prior information concerning the increased aggressive tendencies on the part of the storyteller. This assumption clearly suggests that if the *S* was more aggressive, this tendency should be revealed evenly or equally in all figures within the story. On the other hand, the hero assumption implied that there should be an increase in only three of the four categories. Aggressive acts carried out by the hero against others should increase as a result of the storyteller's increased extrapunitive tendencies. There should also be an increase of aggressive acts carried out by the hero against himself in view of the increased guilt or intrapunitiveness of the storyteller. Finally, there should be an increase in aggressive acts carried out by others against the hero as a result of his perception of the other members of the group as hostile toward him. There was nothing in our analysis of the frustration situation to suggest that the storyteller saw the persons around him as being hostile or aggressive toward each other so we would not predict any change in hostility between non-hero figures. Consequently, we find that the two assumptions agree in predicting significant changes in three of the four categories but are differentiated in their predictions concerning the fourth category.

The reasoning above is readily defensible on rational grounds. The important feature of this derivation, however, is not its invulnerability to logical assault but rather the fact that it was executed prior to analysis of any data except that having to do with a category where no difference was predicted (aggressive acts carried out by heroes against others). In other words, before the fact, the two assumptions, coupled with our detailed knowledge of the frustration situation, seemed to lead to a differentiated prediction which we set out to test. After the fact, there is little doubt that with reasonable motivation and ingenuity either assumption could be rationalized by a sophisticated observer with these or almost any other set of empirical findings.

Procedure

The *S*s in this investigation were 40 male undergraduate students of Harvard University who had been selected so that on the Allport-Kramer Prejudice Scale (1) they fell at the extremes of a group of 575 students enrolled in an undergraduate psychology course. This division of the *S*s into high and low prejudice groups is of no interest in the present inquiry, and our current analysis overlooks the dimension of prejudice except for the fact that experimental and control *S*s were individually matched in terms of prejudice score. The 20 control *S*s were also matched with the experimental *S*s in age. The *S*s were told merely that they were participating in a study of personality structure and development. At the very outset of the study, each *S* was administered individually a shortened version of the Thematic Apperception Test consisting of Cards 3BM, 8BM, 16, and 20. At the end of approximately two months, the 20 experimental *S*s were exposed to an experimentally contrived frustration situation, and immediately following this, they were again given the TAT. The instructions were to make no effort to recall their original story, but if they thought of it first, to put it aside and tell the next story that came to mind. The control *S*s were given the TAT under the same conditions except that there was no intervening frustration experience.

The frustration experience has been fully described elsewhere (9), and it is necessary here only to point out that the experience was carefully divorced from the administration of the TAT and that the *S* was exposed to multiple frustration involving both psychological and complex social motives. The frustration of the latter motives was effected in a group experiment conducted with four confederates, two male and two female, who were ostensibly fellow participants in the study. By disguised manipulation, the experimental *S* was made to fail repeatedly on the group task, which was presumably related to intelligence; he thereby failed to achieve a sizeable financial reward offered for successful performance, and he also kept the other members of the group from winning financial rewards. At various times beginning immediately after the frustration situation, detailed subjective reports were collected, and these, in addition to objective observations made during the conduct of the experiment, enabled us to describe quite accurately just how the *S*s experienced or reacted to this situation.

The TAT protocols were scored simply by counting aggressive acts and then coding them in terms of whether or not they were carried out by the hero of the story and whether or not they were directed toward the hero. The effects of the frustration situation were measured by means of subtracting for each experimental *S* the number of aggressive acts within each category before frustration from the number of such acts after frustration. From this difference score, we then subtracted the difference between the first and second administration of the test for the matched control *S*. In other words, the scores that we are concerned with represent the difference between the first and second test administrations for the experimental *S*s, corrected by the equivalent shifts shown by the control *S*s.

Results

The results of our analysis are summarized in Table 5. Utilizing either assumption, we predicted that aggressive acts by the hero against others, by hero against self, and by others against hero, would increase following frustration. There is confirmatory evidence for all of these predictions. The shift is in the predicted direction in all three cases and is statistically significant at the conventional .05 level for acts involving the hero against others, and others against hero. It is just short of this significance level for acts involving the hero against self.

TABLE 5

CHANGES FOLLOWING FRUSTRATION IN INCIDENCE OF VARIOUS TYPES OF AGGRESSIVE ACTS

Shift in Score	Hero against Other	Hero against Self	Other against Hero	Other against Other
Increase				
4			xxx	
3	x	x	xxx	x
2	xxxxxx	xxx	xxxxx	
1	xxxxxx	xxxx	xxxxx	xxxxxx
0	xxx	xxxxxxxx	xxx	xxxxxxx
Decrease				
−1	xxx	xxx	x	xxx
−2		x		xxx
−3	x			
$\bar{X}$	.75	.40	1.75	0
$s_{\bar{x}}$	.33	.28	.29	—
t	2.27	1.43	6.89	—
p	<.02	<.10	<.005	—

When we turn to the fourth column of Table 5, we find the data bearing upon the differential predictions derived from the two assumptions. The hero assumption predicted no change, whereas the other assumption predicted a shift similar to that observed for the categories just discussed. The results are surprisingly definitive as the distribution of change scores has a mean of exactly zero; there is no evidence whatsoever of any shift in this category. In other words, the data we have reported provide strong evidence for the superior predictive efficiency of the hero assumption under the single circumstance where the two assumptions differ in their consequences.

DISCUSSION

It is clear that the results of these two studies provide some warrant for the continued use of the hero assumption. Our findings suggest that under two circumstances the derivations from this crude assumption fit the observed data better than the derivations that can be made from the easy assumption that all figures in the story are equally revealing of storyteller characteristics. Having agreed to this, however, we must hasten to emphasize the importance of research and formulation that will lead us to a more elaborate statement of the hero assumption so that under known conditions we can apply the kind of complexity in analysis that our findings as well as the convictions of most clinicians, imply is necessary in order to derive consistently sensitive and accurate inferences from the instrument.

There is no discussion of the alternatives to the hero assumption nearly so illuminating as Murray's original analysis of the hero distinction where we find the following special cases proposed:

> (1) The identification of subject with character sometimes shifts during the course of the story; there is a *sequence of heroes* (first, second, third, etc.). (2) Two forces of the subject's personality may be represented by two different characters, for example, an antisocial drive by a criminal and conscience by a law-enforcing agent. Here we would speak of an *endopsychic thema* (internal dramatic situation) with two *component heroes*. (3) The subject may tell a story that contains a story, such as one in which the hero observes or hears about events in which another character (for whom he feels some sympathy) is leadingly involved. Here we would speak of a *primary* and a *secondary* hero. Then (4), the subject may identify with a character of the opposite sex and express a part of his personality just as

> well in this fashion. (In a man this is commonly a sign of a high feminine component and in a woman of a high masculine component.) Finally, there may be no discernible single hero; either (5) heroship is divided among a number of equally significant, equally differentiated *partial heroes* (e.g., a group of people); or (6) the chief character (hero in the literary sense) obviously belongs to the subject-object situation; he is not a component of the storyteller's personality but an element of his environment. The subject, in other words, has not identified with the principal character to the slightest extent but has observed him as he would a stranger or disliked person with whom he had to deal. The subject himself is not represented, or is represented by a minor character (hero in our sense) (12, p. 7).

Having tentatively identified such special cases is an important contribution, but it is equally essential to provide a careful specification of how one goes about identifying actual stories and figures that should be interpreted in the light of each of these cases. This, it seems to us, is a promising and largely untouched area of research.

There are several tacks which such investigation might follow. First, it is possible that one might be able to discover within-story cues that would be helpful in deciding whether to apply the simple hero assumption or some more complex version. Second, there is the ever present likelihood that knowledge of which assumption would be most fruitful will depend upon further information concerning the *S* himself. Thus, the appropriate assumption might vary with the cognitive style, character type, cultural background, or intellectual level of the *S*. Third, depending upon the situational context in which the test is administered, the process of story creation may vary so that different assumptions are warranted. When the test is given in a threatening assessment situation, we might find a different interpretive assumption warranted than when the test is given in a permissive clinical setting, where the *S* is voluntarily seeking assistance. There are, of course, many other types of questions that might be asked concerning this aspect of the interpretive process. However, if we knew something about the relation between the variants of the hero assumption and variation in the nature of the story, in the characteristics of the *S*, and in the situational context, we would be tremendously advanced over our present position.

Granted that further research is indicated, do our findings in their present state provide us with useful information? They do! First, as we have indicated, they give modest support for those clinicians and investigators who have habitually employed the hero assumption in their use of the TAT. Second, they provide negative evidence for the various persons working with the instrument who have attempted to eliminate completely the hero assumption in favor of the other alternative we have considered. One may object to these inferences on the grounds that our findings are by no means definitive. Indeed they are not! But, in the absence of definitive findings, one uses the best evidence one can find, and it seems to us that the present studies fit this specification even if only by default.

There remains the interesting question whether our findings have any implications for the research carried out by the many investigators interested in the quantitative study of the TAT who have not employed the hero assumption. What of investigations such as those by Eron (2), Hartman (3), Henry (4), and McClelland *et al.* (11), where there is typically no distinction between hero and other? In spite of the highly tentative nature of our findings, they do seem to imply that these investigators might have demonstrated somewhat greater test sensitivity if they had recognized the difference between characteristics displayed by hero and by other figures in their efforts to relate test attributes to independent measures. This same possibility exists in connection with some of our own research (8, 10) and poses an intriguing problem for further investigation.

SUMMARY

We have conducted two studies designed to test the comparative effectiveness of the assumption that there is customarily a single figure in TAT stories which is particularly revealing of the storyteller's own attributes as opposed to the assumption that all figures in the stories are equally revealing of the subject's characteristics. In the first study, 30 female undergraduate subjects were asked to judge the similarity between themselves and the figures in TAT stories they had created. They also reported on the similarity between these figures and other persons they had known, and described what had led them to tell each story. The results of the study confirmed our

prediction derived from the hero assumption that hero figures would more often be identified as similar to self or else denied any similarity to self, while non-hero figures would more often be identified as similar to other persons or else would represent general stereotypes.

The second study examined changes in TAT protocols following a frustration experience. Aggressive acts carried out by heroes against others and against the self, and also aggressive acts carried out by others against the hero, all increased following frustration. There was no change in the incidence of aggressive acts carried out by others against others. The hero assumption had predicted just this pattern of results, while the alternative assumption incorrectly predicted consistent increases in all four types of aggressive acts.

Thus, the results of both studies provide evidence supporting the utility of the conventional assumption of a hero in each TAT story. The findings also suggest, however, that under certain conditions a more complex set of assumptions may be desirable or necessary.

REFERENCES

1. ALLPORT, G. W., & KRAMER, B. M. Some roots of prejudice. *J. Psychol.*, 1946, **22**, 9–39.
2. ERON, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64**, No. 9 (Whole No. 315).
3. HARTMAN, A. A. An experimental examination of the Thematic Apperception Technique in clinical diagnosis. *Psychol. Monogr.*, 1949, **63**, No. 8 (Whole No. 303).
4. HENRY, W. E. The Thematic Apperception Technique in the study of culture-personality relations. *Genet. Psychol. Monogr.*, 1947, **35**, 3–135.
5. LINDZEY, G. An experimental examination of the scapegoat theory of prejudice. *J. abnorm. soc. Psychol.*, 1950, **45**, 296–309.
6. LINDZEY, G. Thematic Apperception Test: Interpretive assumptions and related empirical evidence. *Psychol. Bull.*, 1952, **49**, 1–25.
7. LINDZEY, G., & HEINEMANN, SHIRLEY H. Thematic Apperception Test: Individual and group administration. *J. Pers.*, 1955, **24**, 34–55.
8. LINDZEY, G., & NEWBURG, A. S. Thematic Apperception Test: A tentative appraisal of some "signs" of anxiety. *J. consult. Psychol.*, 1954, **18**, 389–395.
9. LINDZEY, G., & RIECKEN, H. W. Inducing frustration in adult subjects. *J. consult. Psychol.*, 1951, **15**, 18–23.
10. LINDZEY, G., & TEJESSY, CHARLOTTE. Thematic Apperception Test: Indices of aggression in relation to measures of overt and covert behavior. *Amer. J. Orthopsychiat.*, 1956, **26**, 567–576.
11. MCCLELLAND, D., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
12. MURRAY, H. A. *Thematic Apperception Test Manual.* Cambridge: Harvard Univer. Press, 1943.
13. PIOTROWSKI, Z. A. The Thematic Apperception Test of a schizophrenic interpreted according to new rules. *Psychoanal. Rev.*,1952, **39**, 230–251.
14. PIOTROWSKI, Z. A. TAT newsletter. *J. proj. Tech.*, 1952, **16**, 512–514.

Received June 19, 1957.

REINFORCEMENT OF AFFECT RESPONSES OF SCHIZOPHRENICS DURING THE CLINICAL INTERVIEW[1]

KURT SALZINGER AND STEPHANIE PISONI

Biometrics Research, New York State Department of Mental Hygiene

BEHAVIOR theory has recently expanded its scope to deal with verbal behavior (6). Greenspoon (1) demonstrated the effectiveness of verbal reinforcers upon a subject's rate of utterance of plural vs. nonplural words. Hildum and Brown (2) showed the effect of verbal reinforcement upon attitude statements. Verplanck (7) used verbal reinforcers during conversations to condition opinion statements. Finally, Salzinger (5) investigated the conditioning process in clinical interviews with schizophrenics.

While these studies have supplied evidence for the validity of the application of reinforcement theory to verbal behavior, a good deal of research is still necessary. The present experiment is designed to study (*a*) reliability of response unit isolation, i.e., to what extent the interviewer can respond reliably with reinforcement to the patient's verbal behavior, (*b*) the effect of different sources of reinforcements (different interviewers) upon the verbal behavior of the interviewee, and (*c*) the relationship between the number of reinforcements and the number of responses in extinction.

Since a patient's ability to express affect is usually evaluated through the interview and is considered an important criterion both for diagnosis and prognosis of schizophrenia, the conditions under which affect is evoked by the interviewer might have theoretical importance for arriving at laws describing interview behavior and practical importance in furnishing an objective method for the evaluation of "flatness" of affect. An attempt was made, therefore, to examine the effect of reinforcement upon schizophrenics' output of affect responses in an interview.

METHOD

Subjects

Twenty-four female and twelve male hospitalized schizophrenics from the age of 18 to 50, with a median of 34.3 years, were selected from the admissions to Brooklyn State Hospital. Patients were classified as schizophrenic upon their current admission to the hospital distribution center. One was later rediagnosed as manic-depressive. Nineteen had been previously hospitalized, and 17 had no history of previous hospitalization.

None of the patients received any somatotherapy such as insulin, electric shock, or tranquillizing drugs for at least one week before the first interview or during their participation in the study.

The first 20 patients interviewed were placed in the experimental group. Fourteen were females and six males, with a median age of 32.0 years and a median number of years of education of 10.3. The other 16 patients were placed in the control group. Twelve were females and 4 males with a median age of 34.5 years and a median number of years of education of 11.5.

Experimental Procedure

All patients were interviewed one week after their arrival in the hospital. The *S*s in the experimental group were interviewed once by a female *E* and once by a male *E* on two consecutive days for a period of 30 minutes each. Eleven of the patients were first interviewed by the male; nine were first interviewed by the female *E*. The *S*s in the control group were interviewed once only, nine by the female and seven by the male *E*. All interviews were recorded with the apparatus in full sight of both patient and interviewer.

The interview was presented to the patients as a routine mental hospital procedure. For the first interview, *E* brought the patient into the experimental room and explained that the interview was being conducted to help him. The second interview was introduced by telling the patient that it is helpful to patients to be interviewed more than once despite the fact that this might mean a repetition of their story. All other interview procedures were the same for the second as for the first interview.

The *E* questioned the patient about the following items: name, age, marital status, children, and siblings. The patients answered these questions with little hesitation, thus making it possible for *E* to begin the interview by repeating the answer given, by writing it down, and by saying such words as "mmm-humm," "uhhuh," "I see," etc. This procedure was adopted in an effort to obtain factual information upon which subsequent

---

[1] This study was supported by research grant M586(C3) from the National Institute of Mental Health of the National Institutes of Health, Public Health Service. The authors are indebted to J. Zubin and S. Sutton for their constant help and interest in this study and to Suzanne Salzinger for her editing. The authors are also grateful to N. Beckenstein, Director of the Brooklyn State Hospital, A. Glen, L. Olinger, L. Granick, and M. Portnoy for their cooperation in helping to make available the sample used for this study. This paper was presented in part at the Eastern Psychological Association, New York City, 1957.

interview questions could be based and to establish *E* as a source of reinforcement, in this way encouraging the patient to speak in the presence of *E*. The main part of the interview was then initiated with the question, "Would you tell me why you are here in this hospital?"

Interviews with the experimental group were conducted in the following manner: During the first 10 minutes (operant level), the base rate of spontaneous affect responses (see definition below) was determined. The *E* asked questions but did not reinforce any statement made by the patient. Reinforcement was defined as *E*'s verbal agreement through the use of such words as "mmm-hum," "I see," "yeah," etc., with statements made by the patient.

During the second 10 minutes (conditioning), *E* continued to question the patient and reinforced each affect response by immediately following each expression of affect with verbal agreement.

During the third 10-minute period (extinction), *E* withheld all reinforcement but continued asking questions.

Interviews with the control group also lasted 30 minutes, during which time *E* asked questions but did not reinforce any of the patient's responses. This procedure was identical with the operant level phase of the experimental group procedure.

Definition of Response

The response class of affect for this experiment was defined as any statement describing or evaluating the state (other than intellectual or physiological) of the patient by himself. The response class therefore included all statements beginning with the pronouns "I" or "we" and followed by an expression of affect. Examples include such expressions as: I am satisfied, I'm happy, We enjoyed it, I like him, I'm very close to him, I was mad at him, We hated her, I'll always be jealous of him, I am upset, I am a lonely person, I was so ashamed, I'm sorry for him, I feel . . . (followed by any words), I was frightened, We couldn't take it, I always suffer, I had a fright, etc.

Quotations in which affect is referred to the speaker, although fitting all other criteria, were excluded on the basis of not being direct expressions of the patient's affect. An example of this was, "My husband said I didn't feel good." Statements like "I am happy and excited" were considered as one affect statement only, since the pronouns "I" or "we" did not precede the second affective word. On the other hand, incomplete (in the sense that the object of the affect is not mentioned) statements like "I love. . ." or "We feared. . ." were viewed as separate responses.

Certain types of private events or internal states were excluded from the response class of affect because they referred primarily to intellectual processes, to actions which are sometimes but not always associated with affect, or to desires which appear to constitute a class of responses different from the affect class as defined here. I am confused, I am confident, I would like to. . ., I want, I was surprised, I am not well, We forgave him, I threaten her constantly, I didn't trust them, etc., are examples.

A count per minute was taken of statements belonging to the general class of self-references (statements beginning with "I" or "We") in order to compare changes in the occurrence of this class with those of the class of self-referred affect statements. In other words, self-referred statements included both self-referred affect statements as well as self-referred nonaffect statements.

Interviewer Questions

After the initial question, "Why are you here?" *E* asked additional questions only when the patient ceased talking for at least two seconds. Some or all of the following topics were discussed during each interview: reasons for being in the hospital and causes for illness; patient's relationships to his parents, siblings, fellow employees, employers, fellow students and teachers, wife or husband, children, friends; patient's activities during free time, and plans for the future. The *E* made an attempt to balance these topics over the different conditions. For instance, if the patient discussed the symptoms of his illness in the operant level condition, *E* asked questions regarding the possible causes of the illness during conditioning and brought the patient back to these topics during extinction. As long as the topics were approximately balanced over the three conditions, however, *E* took his cues as to topic from the content of the patient's statements. Both the number of topics and the order in which they were discussed varied from interview to interview.

Questions[2] asking directly for affect such as, "How did you feel about that?" or "Were you happy?" were not used.

RESULTS

Reliability of Response Unit Isolation

A sample of 15 recorded interviews was coded independently for self-referred affect by the two interviewers. Proportions of agreement based on the number of affect statements counted were computed separately for each condition of each interview and ranged from .79 to 1.00. Examination of the disagreements revealed that they were primarily due to poor recording. It was therefore concluded that the affect responses as defined in this experiment can be objectively isolated and counted.

---

[2] A pilot study of the effect of interviewer questions of this type of interview was undertaken by the following four graduate students: Ruth Beach, R. S. Feldman, P. Goldberg, and Marilyn G. Hamlin. They found no significant relationship between the conditioning effect and the following: specific vs. nonspecific questions, the introduction of new topics vs. the continuation of old topics, the total number of questions asked, and the number of questions indirectly leading to affect (e.g., How did you react to that?). A larger more definitive study of question effects is presently being undertaken by the authors.

Interviewer Differences

The adequacy of the definition of the response was examined by having the two *E*s who served as interviewers in the experiment independently code the affect responses of the same 15 recorded interviews. This procedure made it possible to test whether both *E*s would have reinforced the same responses in identical interviews.

In order to determine whether the two interviewers evoked a different number of affect statements, comparisons were made between the two interviewers on each condition of the initial interviews of the experimental group and the control group interviews. The Mann-Whitney test yielded no statistically significant differences ($p > .05$), suggesting that the two interviewers evoked approximately the same number of affect statements in their respective interviews. The exact p levels for the experimental group interviews were .79 for the operant level, .54 for the conditioning period, and .52 for the extinction period. For the control interviews, the p levels were considerably lower, although still not significant. The p level for the first 10-minute period was .07, for the second .06, and for the third .22.

Base Level of Affect

In order to determine whether the experimental and control groups differed initially in the amount of affect spontaneously emitted, a statistical comparison of the number of affect statements given in the first 10 minutes of the control interviews with the number of such statements given during the operant level of the experimental interviews was made. The difference was not statistically significant ($p = .37$) by the Mann-Whitney test (4).

Conditioning Effect

The difference between operant level, conditioning, and extinction for the initial interviews of the experimental group was tested by Wilcoxon's nonparametric analysis of variance (8) and found to be statistically significant ($p < .01$). The greatest number of affect statements was emitted during conditioning (sum of ranks = 51.0), the next greatest during the operant level (sum of ranks = 39.5), and the least during extinction (sum of ranks = 29.5). The difference between the three conditions of the second interviews of the experimental group was not statistically significant ($.2 > p > .1$). Inspection of the sums of ranks for the three conditions, however, still revealed the greatest number of affect statements in the conditioning period (sum of ranks = 45.0), the next greatest number of affect statements in extinction (sum of ranks = 41.5), and the smallest number during operant level (sum of ranks = 33.5). The fact that the second interview did not yield a statistically significant difference between the conditions appears to be due largely to the greater number of responses emitted during extinction in the second interview in contrast to the first. This is also evident in the comparison of Figs. 1 and 2, where responses during extinction were plotted as a function of number of reinforcements, from which one can see the steeper slope in the second than in the first interview. In other words, this result does not indicate a lack of reliability over time but, rather, that apparently fewer reinforcements were necessary for the same number of extinction responses in the second than in the initial interview. This effect is generally reported for reconditioning.

When the same test was used to compare the three 10-minute periods of control group interviews, no significant difference was found ($p = .75$). The greatest number of affect responses was emitted during the last 10 minutes of the interview (sum of ranks = 36.0), the next greatest during the second 10 minutes (sum of ranks = 34.0), and the smallest number during the first 10 minutes (sum of ranks = 32.0).

The Mann-Whitney test was used to compare each 10-minute period of the experimental group interviews to its corresponding period in the control interviews. Comparison of the first 10 minutes of the experimental with the control interviews yielded no significant difference ($p = .37$). Comparison of the last 10 minutes also yielded no significant difference ($p = .92$). The difference between the second 10-minute period of the control-group interviews and the conditioning period of the experimental-group interviews was statistically significant ($p = .03$), using the one-tailed test hypothesis that the experimental group would emit more affect than the control group.

Figs. 3 and 4 represent individual cumulative curves of affect responses over the three conditions of the experimental group interviews. Fig. 3 shows the curves of three individuals whose rate of response was modified by reinforcement.

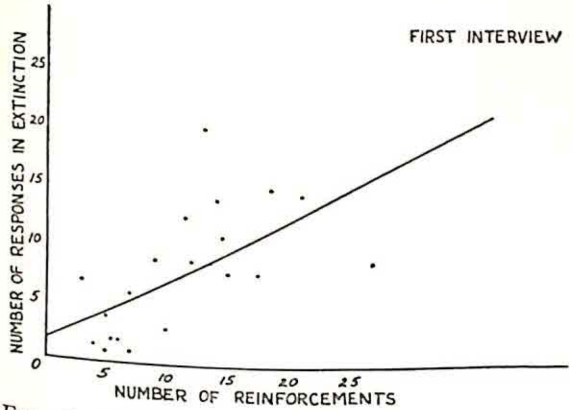

FIG. 1. Number of Affect Responses During Extinction as a Function of Number of Reinforcements in the First Interview

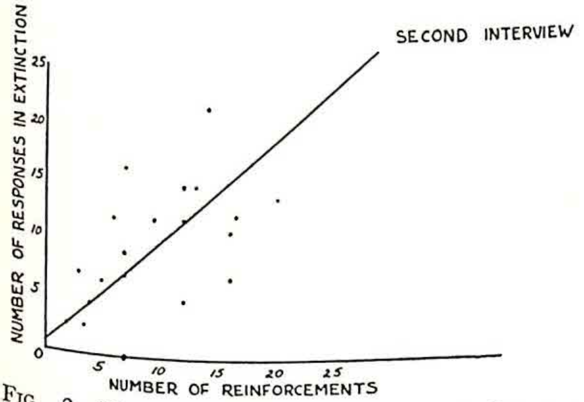

FIG. 2. Number of Affect Responses During Extinction as a Function of Number of Reinforcements in the Second Interview

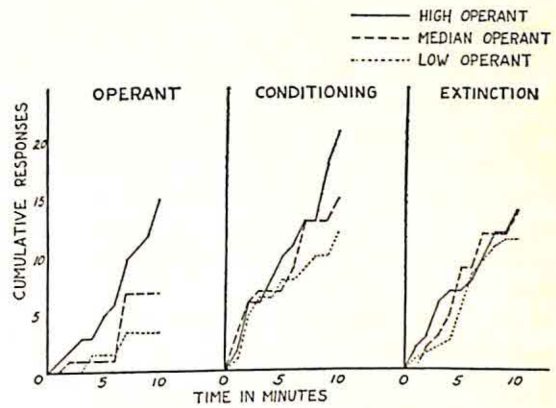

FIG. 3. Individual Cumulative Response Curves for Three Patients Who Showed the Conditioning Effect

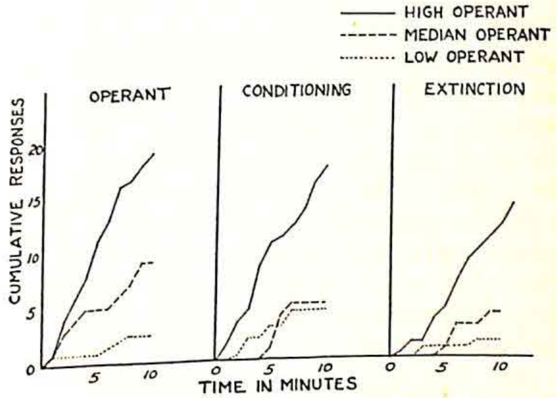

FIG. 4. Individual Cumulative Response Curves for Three Patients Who Showed Little or No Evidence for Conditioning

These individuals were selected on the basis of being representative of low, medium, and high rates of response in operant level. Fig. 4 shows the curves of three individuals whose rates of response were not modified by reinforcement. These individuals were also representative of their group.

In order to gauge the lawfulness of the process of conditioning in the experimental group, the number of responses of each patient during extinction was plotted against the number of reinforcements in conditioning. It was found that the relationship could be described by a linear equation, i.e., the greater the number of reinforcements administered, the greater the number of responses emitted during extinction. The goodness of fit can be seen by examining Figs. 1 and 2. Two patients appear to deviate markedly from the rest of the sample. The diagnosis of one of these patients was changed from schizo-affective to manic-depressive psychosis. She received 13 reinforcements and gave 20 extinction responses. The other deviate from the group, who received 27 reinforcements and gave only 9 extinction responses, was later found to be hard of hearing.

In order to investigate further the effects of reinforcement, rank-order correlation coefficients were computed between every pair of conditions, separately for the first and second experimental interviews and the control interviews. As expected for the experimental group, the highest correlations were found between the number of reinforcements and the number of extinction responses for both the first (.73) and second (.60) interviews. The correlations between all other pairs of conditions were much lower, varying within a restricted range from .41 to .50 (see Table 1). The correlation between the two extinction periods of Interviews 1 and 2 was .41 ($p < .05$) and that between extinction of Interview 1 and [illegible] Interview 2 was [illegible]

TABLE 1

RANK-ORDER CORRELATIONS BETWEEN THE THREE TEN-MINUTE PERIODS OF THE INTERVIEWS

Conditions	Experimental		Control
	First Interview	Second Interview	
Conditioning vs. Extinction (2nd vs. 3rd 10 minutes)	.73**	.60**	.74**
Operant level vs. Conditioning (1st vs. 2nd 10 minutes)	.46*	.47*	.85**
Operant level vs. Extinction (1st vs. 3rd 10 minutes)	.41*	.50*	.70**

* $p < .05$.
** $p < .01$.

In direct contrast to the results of the experimental group, the correlations between the conditions in the control group were all evenly high, ranging from .70 to .85.

Since every affect statement in the second period of the experimental group was reinforced, the question arises whether the correlation between number of reinforcements and number of responses in extinction merely reflects a correlation between the affect statements in different parts of the interview.

Kendall's tau (3) was computed in order to partial out the correlations between operant level and conditioning and operant level and extinction from the reinforcement-extinction correlation in the experimental group. The tau of .58 ($p = .0002$) between number of reinforcements and extinction responses for the first interview became .53, and the tau of .45 ($p = .002$) for the second interview became .37. When the first 10-minute period of the control interviews was partialed out of the correlation between the second and third 10-minute periods, the tau of .59 ($p = .002$) became .38, a drop of .21, whereas the corresponding drop in the experimental group was only .05.

This indicates that only a small part of the correlation between number of reinforcements and number of responses in extinction can be accounted for by the correlation between operant level and number of reinforcements; the correlation between the second and third parts of the control group, on the other hand, can be accounted for in larger part by the correlation between the first and second part of the interview.

The rate of making self-referred statements was found to be invariant over the three conditions of the first and second interviews of the experimental group as well as the interviews of the control group. This was tested by Wilcoxon's (8) nonparametric analysis of variance ($p > .05$). Inspection of the sums of ranks for the first experimental interview (based on an N of 20) indicated a trend toward decreasing frequency of self-referred statements from operant level (sum of ranks = 48.0) to conditioning (sum of ranks = 37.0) to extinction (sum of ranks = 35.0), while in the second interview, the greatest number of self-referred statements appeared in conditioning (sum of ranks = 46.0), the next greatest in extinction (sum of ranks = 39.0), and the smallest in operant level (sum of ranks = 35.0). In the control group (based on an N of 16), the greatest number of self-referred statements was emitted during the first 10 minutes (sum of ranks = 38.5), the next greatest during the third 10 minutes (sum of ranks = 3.5), and the smallest during the second 10 minutes (sum of ranks = 30.0).

DISCUSSION

While it is true that two different interviewers evoked the same number of affect statements within the margin of random error, the fact that two of the comparisons approached the .05 level of significance would seem to indicate that there is still room for greater control of interviewer behavior or that such interviewer characteristics as sex, age, appearance, etc., play an important role in controlling the interviewee's behavior.

While Verplanck (7) was unable to test the constancy of the response used during any one conversation or among different conversations, the recording done in this experiment made it possible to demonstrate that a response class decided upon prior to the interview can be reliably reacted to by different interviewers and coders.

Although this experiment gave definite evidence for conditioning, there was much variability among individuals. This was not surprising in view of the fact that the interviewing was carried out with schizophrenics and the response conditioned one of affect statements. This variability, both in operant level and in susceptibility to reinforcement, might well provide an objective prognostic measure of degree of flatness of affect. Such a measure would be of value since there are strong indications (10) that marked flatness of affect augurs badly for the outcome of schizophrenia. Follow-up of the patients interviewed in this sample may allow an exact test of the relationship between outcome of illness and flatness of affect.

Perhaps one of the most interesting findings of this study was the linear relationship between number of reinforcements and number of responses in extinction. In a similar study, Williams (9) found that the relationship between number of reinforcements and number of responses during extinction for food-deprived rats was also linear up to about 30 reinforcements. Since fewer reinforcements were administered in the present study, it will certainly be of interest to try to duplicate the rest of Williams' curve for verbal behavior. The results become even more dramatic when account is taken of the fact that flatness of affect, as defined here by frequency of affect statements, appears to vary directly as a function of the interviewer's reinforcing behavior. The implications for the regular psychiatric interview are self evident.

The two patients in the experimental group who showed atypical relationships between reinforcement and extinction, are noteworthy because both deviated in directions that seem sensible on a *post hoc* basis. The hard-of-hearing individual got more reinforcements for the number of responses she emitted in extinction than the *S*s in the rest of the sample. This, of course, is exactly what one might expect if *S* could not hear all the verbal reinforcements given her. The patient whose diagnosis was changed from schizophrenic to manic-depressive psychosis, manic stage, gave many more affect statements in extinction than might be expected from the number of reinforcements she received. This observation suggests the possibility that if a sample of manic-depressives was administered the same number of reinforcements as the schizophrenic group, a linear relationship might similarly be found but with a steeper slope. Manic-depressives may require a smaller number of reinforcements than schizophrenics for the same number of responses in extinction.

The correlation coefficients between all possible combinations of conditions were computed in an attempt to see whether the relationship between number of reinforcements and number of responses during extinction could be explained simply as a correlation that might be obtained between any two 10-minute periods of the same interviews. Table 1 shows that while correlations occurred between all conditions, the highest were between the number of reinforcements and number of responses during extinction in the experimental group. Furthermore, upon partialing out the operant level correlation to study the relationship to be expected between any two 10-minute periods, no substantial change occurred in the correlation between number of reinforcements and frequency of response in extinction. The correlations between the conditions in the control group yield further evidence for the conditioning effect in the experimental group. The fact that they are all approximately equally high, whereas the reinforcement-extinction correlations in the experimental group are outstandingly high by comparison to the operant-extinction and operant-reinforcement correlations, argues strongly for the effect of reinforcement.

A final control on the effect of reinforcement in the interview, that of the rate of self-referred statements, was used in order to check on the possibility that the reinforcement functioned merely to make the patient feel more at ease and therefore talk more about himself. This analysis indicated that the conditioning effect was specific in producing an increase in self-referred affect statements and not in increasing the general class of self-referred statements.

SUMMARY

Thirty-six hospitalized schizophrenics were included in this study. Twenty of them (the experimental group) were interviewed for a period of 30 minutes each on two consecutive days by two interviewers. The other 16 (the control group) were given one interview only, which lasted for 30 minutes. Each interview in the experimental group consisted of a 10-minute operant level, during which *E* only asked questions necessary to keep up the patient's talk but did not respond to the patient's speech; 10 minutes of conditioning, during which *E* reinforced by agreement all self-referred affect statements; and, finally, 10 minutes of extinction, during which *E* with-

held all further reinforcement. Each interview in the control group consisted of 30 minutes of operant level only.

It was demonstrated that a difference in interviewers or sources of reinforcement per se need not produce discrepant results during an interview when utilizing a standard procedure for interviewing. It was further shown that a verbal response class can be reliably isolated and reacted to. Conditioning of the response class of self-referred affect statements was found to be possible with schizophrenics during an otherwise usual clinical interview. The relationship between number of reinforcements and number of responses in extinction was described by means of a straight line, i.e., the greater the number of reinforcements, the greater the number of extinction responses.

The lawfulness of these findings indicates that the clinical interview is subject to investigation by experimental techniques. Furthermore, a controlled interview may prove useful as a research tool.

REFERENCES

1. GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.
2. HILDUM, D. C., & BROWN, R. W. Verbal reinforcement and interviewer bias. *J. abnorm. soc. Psychol.*, 1956, **53**, 108–111.
3. KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.
4. MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Stat.*, 1947, **18**, 50–60.
5. SALZINGER, SUZANNE. Rate of affect response in schizophrenics as a function of three types of interviewer verbal behavior. Paper read at East. Psychol. Ass., Atlantic City, March, 1956.
6. SKINNER, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.
7. VERPLANCK, W. S. The control of the content of conversation: Reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, **51**, 668–676.
8. WILCOXON, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid, 1949.
9. WILLIAMS, S. B. Resistance to extinction as a function of the number of reinforcements. *J. exp. Psychol.*, 1938, **23**, 506–521.
10. ZUBIN, J. Role of prognostic indicators in the evaluation of therapy. Paper read at Conference on the Evaluation of Pharmacotherapy in Mental Illness, Washington, D. C., 1956.

Received June 28, 1957.

THE EFFECTS OF PUNISHMENT (ELECTRIC SHOCK) ON PERCEPTUAL LEARNING[1]

HAROLD J. McNAMARA, CHARLES M. SOLLEY, AND JOHN LONG

The Menninger Foundation

THE purpose of these studies was to explore the consequences of the association of punishment with percepts and to trace the residual effects of such associations in several aspects of perceptual and memoric organization. Past experiments in this area (e.g., 1, 4, 7, 10, and 11) have not yielded consistent results on the questions, "What are the conditions of learning which lead to 'perceptual emphasis' and what are the conditions of learning which lead to 'perceptual deemphasis' of percepts associated with punishment." Although Tolman (12) speaks of punished events standing out as things to be avoided, and Postman and Brown (6) have shown that people can become sensitized to objects of failure, there are strong data indicating that, under certain conditions of learning, punished events are "perceptually repressed" or "perceptually deemphasized" (4, 7, 10). The use of radically different experimental procedures and tests, different stimulus materials, different intensities of punishments, etc., make interexperimental comparisons almost impossible. The results of such studies, however, suggest several important variables, and a series of experiments was designed in which these pertinent variables were manipulated within a single procedural framework.

In previous research, the time interval between the percept and the punishing event varied unsystematically from study to study. Smith and Hochberg (11) used contiguous pairing; Pustell (7) used .2 second; Dulany (4) used 2 seconds; and Ayllon and Sommer (1) used 3 seconds. Since research on classical conditioning indicates that the amount of associative strength is a function of the CS-US interval, it is necessary to vary this time interval systematically or else to hold it constant.

Still another factor is the intensity of the punishing agent, which varies considerably from study to study and is reported in terms of subjective evaluation, e.g., "unbearable" (4, 7), "quite painful" (10), etc., as well as in terms of objective voltage levels (11). It was decided to use both subjective evaluations and objective criteria, and to determine the relationship between these two methods of ordering "intensity" of punishment.

The third major variable that could play an important role in the association of punishment with percepts is the possibility of escape from the punishing agent. Ayllon and Sommer (1) had permitted subjects to escape from the punishment; Pustell (7) and Dulany (4) used a design that permitted neither escape nor avoidance. Reece's study (10) involved both escape and no-escape conditions, and he found striking differences between these two conditions both in terms of threshold changes for the associated percept and in terms of subjective interpretation of the "meaning" of the electric shock.

EXPERIMENT I

Each experiment in the series followed the same basic paradigm, viz., that employed by Ayllon and Sommer (1). The first experiment was designed to obtain information about the effects of electric shock at two levels of intensity, 40 v and 60 v, under restrictions of no-escape and simultaneous association. It was hypothesized that *S*s who rated the shock "moderate" to "very unpleasant" would perceive the shock-associated stimulus in the test situation, whereas those *S*s who rated the shock "not unpleasant" to "slightly unpleasant" would perceive the nonshocked stimulus in the test. This hypothesis was based on the results of the Ayllon and Sommer study (1).

Method

Apparatus. The stimulus materials used in the present study were the same three-dimensional reversible profiles used in the Ayllon-Sommer study (1). These are shown in Fig. 1. The profiles were formed by grooves in 12″ × 12″ plaster casts. The grooves were

---

[1] This research was supported by a research grant (M-715C) from the National Institute of Mental Health of the National Institutes of Health, US Public Health Service. The authors are indebted to Gardner Murphy who gave invaluable guidance and suggestions and to Robert Sommer who provided much of the original stimulation for these studies.

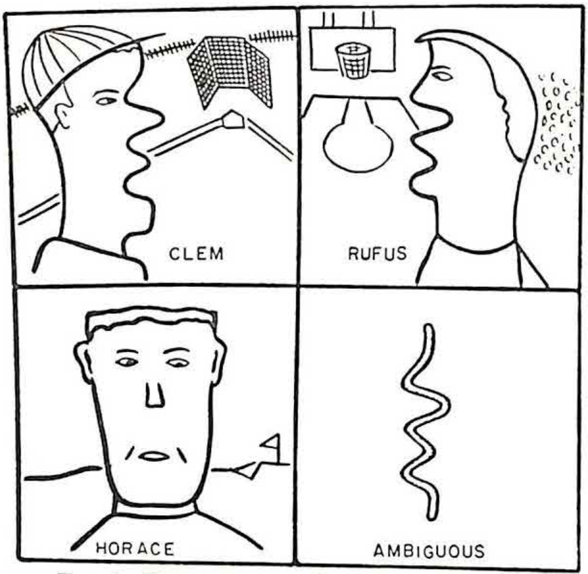

Fig. 1. Diagrammatic Drawings of Plaques

8″ in length, ⅝″ in width, and ⅛″ deep. There were four of these plaques: Rufus, a left-pointing face; Clem, a right-pointing face; Horace, a full face used as a set-breaking figure; and a posttest profile line similar to both Rufus and Clem. Since the right-pointing, left-pointing, and ambiguous posttest profiles were cast from the same mold, the center profile line was identical for each figure. Meaningful backgrounds were painted in colored enamel on the plaques so that during the tracing of Rufus or Clem, visual recognition of the other face would be precluded. *S* wore a rubber glove on the tracing hand to minimize extraneous tactual cues and compel him to rely solely on his figure-ground organization of the center line in determining the identity of the particular face.

The shock apparatus used in the present study was a Variac type with an adjustable time delay system. With this apparatus, the delay and duration of shock could be controlled precisely. The variac was equipped with an attached voltmeter that made it possible to compensate for fluctuations in line voltage. Shock was controlled by a foot pedal that made it difficult for *S* to predict shock by means of body movement cues from *E*.

Shock was applied to the index and middle finger of the left hand. The electrodes were rubber cups filled with moist 10% zinc oxide compound, mounted on a small wooden block and covered with a tight elastic band (glove-like) from which *S* could not remove his hand. In this way, contact pressure, contact area, and moisture from perspiration were controlled and held constant for each *S*.

Subjects. The *S*s in Experiment I were 29 female undergraduate students enrolled in the elementary psychology course at the University of Kansas. None of these *S*s knew, prior to the experiment, that shock would be used. Fifteen *S*s were assigned to the Clem—Shock, Rufus—No-shock group, and 14 were assigned to the Clem—No-shock, Rufus—Shock group.

Procedure. The method was similar to the Ayllon-Sommer procedure with some changes. The Ayllon-Sommer procedure was to present the profile that was to be associated with shock, administer an electric shock of one-second duration three seconds after completion of the tracing, remove the profile, and then present the next profile. The major change from this procedure was to administer the shock *during* the tracing of the plaque (specifically, when *S*'s index finger made contact with the mouth of the profile).

In addition to this change, several minor procedural modifications were incorporated into the present experiment. The order of presentation of the faces was changed in the hope of overcoming the preference Ayllon and Sommer had found for Rufus, the left-pointing profile.

The order of presentation of the faces was as follows: Clem, Horace, Horace, Rufus, C**, H, R*, C, R, H, R*, C**, C, H, R, R*, H, C**, R*, H, C**, H, R*, R, C, C**, H. The *E* presented one face at a time and called out the appropriate name while placing the face in front of *S*. Certain of these trials involved electric shock. Those marked with a single asterisk indicate shock for the Rufus—Shock group; a double asterisk indicates shock for the Clem—Shock group. In order to deal with the criticism that shock during the tracing may have a distracting effect, the faces were traced twice and shock was always given on the *second* tracing of the face.

Another method was used to define the severity of shock along with ratings of the shock by *E* and by *S*. Shock was defined in terms of the applied voltage, that is, half of the *S*s were given 40-volt shock, and half were given 60-volt shock. Experience of the shock as to degree of unpleasantness was determined by *S*'s rating of the shock along a five-point scale from "not-unpleasant" to "very unpleasant," and by *E*'s rating of *S*'s behavioral reactions as either slight, moderate, or strong. The *E*'s rating of shock was made after blindfolding *S* and prior to *S*'s first response in the test trials.

E rated each *S* in terms of the overt reaction shown to the shock: strong, moderate, and slight. The criteria for rating a reaction as *strong* were such responses as flinging of the arm whenever shock was applied, raising of finger from the table in absence of shock, and verbalizations regarding the shock (e.g., "Gee, that hurts," "How long is *this* going to last"). If *S* showed any or all of these reactions to a lesser degree, *E* rated her reaction to the shock as moderate. On the other hand, if *S* showed few or none of the above mentioned behaviors, *E* rated her reaction to the shock as *slight*. Immediately following the test, *S* was asked to rate the severity of the shock on a five-point scale from "not unpleasant" to "very unpleasant."

In the test series, there was a total of 27 presentations: 14 of the ambiguous profile line and 13 of Horace (set-breaking full face). The latter was alternated with the ambiguous profile in the following sequence: Ambiguous, Horace, H, H, A, A, H, A, H, H, A, H, A, H, A, A, H, A, H, H, A, H, A, A, H, A, A.

Results

Analysis of the data based on *S*'s first response to the ambiguous figure is presented in Table 1. The data show that of the entire group, 21 *S*s reported the nonpunished face as their first response while 8 *S*s reported

TABLE 1

FIRST RESPONSE GIVEN TO AMBIGUOUS PLAQUE DURING TESTING SEQUENCE (EXP. I)

Measure of Shock	Intensity	First Response		p** (Two-tail)	
		Punished Face	Nonpunished Face		
Physical	40 volt	6	9	N.S.	N.S.
	60 volt	2	12	.01	
Rating by S	Slight	6	10	N.S.	N.S.
	Moderate	2	11	.02	
Rating by E	Slight	7	9	N.S.	.05
	Moderate	0	12	.01	
Rating by S and E*	Slight	4	7	N.S.	N.S.
	Moderate	0	9	.01	

* Instances ($N = 9$) where S and E disagreed were discarded.

** χ^2 corrected for continuity.

TABLE 2

SUMMARY OF ANALYSES OF MEAN NUMBER OF PUNISHED FACES REPORTED TO AMBIGUOUS FIGURE (EXP. I)

Measure of Shock	Intensity	$\bar{X}$	N	$SD_{\bar{X}}$	t tests	
					Inter	Intra
Physical	40 volt	7.80	15	.65	3.08**	1.23
	60 volt	5.43	14	.44		3.46**
Rating by S	Slight	7.25	16	.65	2.80**	.38
	Moderate	5.92	13	.78		1.80
Rating by E	Slight	7.56	16	.65	2.95**	.89
	Moderate	5.50	12	.48		2.70***
Rating by S and E*	Slight	7.82	11	.90	2.56***	.95
	Moderate	5.11	9	.71		2.79**

* Instances ($N = 9$) where S and E disagreed were discarded.

** p (Two-tail) .05.

*** p (Two-tail) .01.

TABLE 3

FACE REMEMBERED MOST VIVIDLY (EXP. I) AFTER TRAINING AND TEST

Measure of Shock	Intensity	Remembered most vividly		p (Two-tail)	
		Punished Face	Nonpunished Face		
Physical	40 volt	10	2	.05	.05
	60 volt	6	8	N.S.	
Rating by S	Slight	10	3	.10	.10
	Moderate	3	6	N.S.	
Rating by E	Slight	11	3	.05	.10
	Moderate	5	7	N.S.	
Rating by S and E*	Slight	8	2	.10	.10
	Moderate	3	6	N.S.	

* Instances ($N = 9$) where S and E disagreed were discarded.

the punished face ($p < .01$). There was no significant difference found between groups. However, all measures of severity of shock reveal that the less intense shock had a less marked effect than did the more intense shock.

Analysis of the mean number of total responses to the ambiguous profile is given in Table 2. Here the difference between groups is significant for all measures of shock, showing that the group receiving or experiencing the stronger shock reported more nonpunished faces than the other group. Intra-group comparisons with chance expectancy are also shown in Table 2. The results are similar to the first response data. The 40-v shock had little or no effect, although there is a trend toward more punished responses; while the 60-v shock group shows a significant reporting of the nonpunished face.

Immediately after the test series, the *S*s were asked which face they remembered most vividly. If they answered "Horace," the set-breaking profile, they were asked "which next?" Table 3 presents the faces, other than Horace, that were remembered most vividly. Although there is a definite trend for the slight-shock group to remember the punished face, only the physical measure of shock gives significant results. There is also a slight trend for the moderate shock group to report the nonpunished face. However, this tendency is not significant.

EXPERIMENT II

Since the strongest effects had been found using 60 v in Experiment I, a second study was designed to investigate the temporal interval between percept and punishment. In this experiment, the procedure was the same as in Experiment I, except that the 60-v shock was administered three seconds *after* tracing of the plaques.

Subjects

Forty-one female undergraduates at the University of Kansas were used. Twenty-six of these were randomly assigned to the experimental group, and 15 were randomly assigned to the control group.

Results

In this experiment, the primary interest was the investigation of the effect of moderate to very unpleasant shock applied three seconds *after* the tracing of the faces. At this intensity (60 v), only six of the 26 *S*s reported the shock as slightly unpleasant, and 20 *S*s reported the shock as moderately to very unpleasant. In addition, *E* rated 21 of the 26 *S*s as reacting strongly to the shock.

The data on first response and vividness were not significant. The analyses of the total responses to the ambiguous profile for all *S*s, for *S*s who rated the shock as moderate, and for *S*s whom *E* rated as reacting strongly yielded essentially the same result. The mean number of punished responses was 5.54, and the mean number of nonpunished responses was 8.46. A *t* test between these two means yielded a value of 2.80, significant at the .01 level. The results are in accordance with those of Smith and Hochberg (11) but show no trends in the direction reported by Ayllon and Sommer (1).

The temporal relationship between the shock and the stimulus (Ayllon and Sommer gave shock three seconds after presentation of the stimulus) was not found to be a significant variable. In Experiment I, shock during the presentation of the stimulus, and in the present study, shock after presentation gave the same results; *S*s reported a significant number of *nonshocked* faces when strong shock, or shock rated as moderate to very unpleasant, was used.

There is some reason to propose a "repression" model to explain these results. Tolman (private communication) suggested that the effects of punishment upon learning or perception do not evidence themselves immediately; retesting after a time lapse might reveal more pronounced repressive effects. Rapaport (8) makes the similar suggestion that repression occurs *after* a time lapse rather than immediately. For a preliminary test of this proposition, all *S*s in Experiment I were phoned three weeks to two months after the initial testing session. Of 29 *S*s, only 17 *S*s were able to recall either Rufus or Clem. The recall data from these 17 *S*s were compared with their immediate recall at the end of the initial test session, and it was found that four *S*s changed in terms of the face remembered most vividly. All four changed from remembering the punished face to remembering the nonpunished face most vividly. No *S*s switched from the nonpunished to the punished face in delayed recall. These results in conjunction with the data in Experiment I and Experiment II suggest that "repression" is operating in this experimental situation.

Experiment III

On the basis of the first two exploratory experiments, a final study of broader scope was planned, involving the variables of "escape possibilities," physical intensity of shock, and perceived intensity of shock. Obviously, the latter two variables are somewhat correlated.

It was hypothesized, first, that when *S* can escape from shock by alerting himself to certain percepts and not to others, he tends to perceive the shock-associated percept in the test situation. On the other hand, when *S* cannot escape from the shock, "perceptual repression" occurs in that he tends to perceive the nonshocked stimulus in the test situation. The only restriction placed on this hypothesis was that under low intensities of shock (far below pain threshold) there should be a dynamogenic effect (9) and both "escape" and "nonescape" groups should perceive the shock-associated profile in the test.

As the intensity of the shock is increased, second, a progressively greater difference between escape and no-escape conditions is hypothesized: specifically, a significant interaction between intensity of shock and escape conditions.

Our third major hypothesis asserts that perceptual alerting and perceptual repression are a complex function of the physical intensity of the electric shock. We did not attempt to specify the exact form of this functional relationship.

Fourth, we expected a significant correlation between *perceived* intensity of shock and perceptual alerting or perceptual repression.

A fifth hypothesis was proposed tentatively. We had observed in Experiments I and II that there were shifts in recalled vividness of faces with *S*s who perceived the shock-associated profile in the test, whereas *S*s who perceived the nonshock-associated profile in the test recalled the same profile most vividly. We hypothesized on this basis that *S*s brought back into the same experimental room and given a *replication* of the test without repeated training would perceive the nonshocked profile predominantly. The basis for this argument was that "anxiety" associated with the training would continue incubating (Bindra and Cameron [2]) and would subsequently "repress" the shock-associated percept.

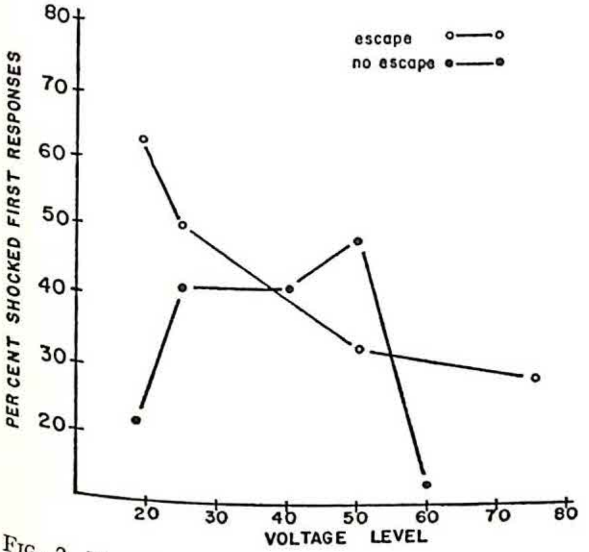

FIG. 2. DISTRIBUTION OF FIRST RESPONSES FOR THE ESCAPE AND NO-ESCAPE CONDITIONS

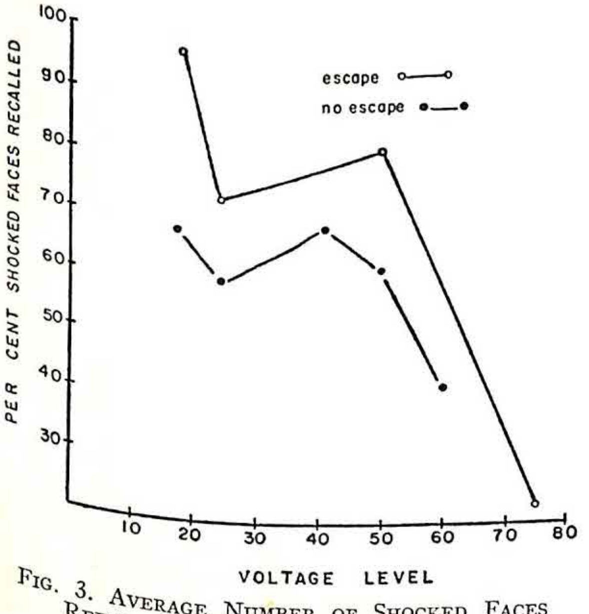

FIG. 3. AVERAGE NUMBER OF SHOCKED FACES REPORTED BY EACH ESCAPE CONDITION

Subjects

Seventy-nine female undergraduates at the University of Kansas served as *S*s. None knew that shock was to be used, and none reported knowing about the study in advance. *S*s were randomly assigned to one of seven experimental conditions.

Apparatus and Procedure

The apparatus and materials were the same as those used in Experiments I and II, and the procedure was basically the same. Shock levels of 18, 25, 50, and 75 volts were used. In the 75-volt condition, only six volunteer *S*s were tested and only under conditions of escape. In the no-escape condition, a tight elastic glove was fitted around the *S*'s hand so that she could not remove her hand from the electrodes. In the escape condition, the elastic glove was omitted, and *S* could remove her hand from the shock apparatus.

For the delayed testing procedure, only the 50-volt group was brought back and given a replication of the test. These *S*s were given only the test procedure and no additional shock was administered. The 30 *S*s were divided into three groups of 10 *S*s with an equal distribution of *S*s from each escape condition. Group I was retested after 7, 14, and 28 days; Group II after 14 and 28 days; Group III after 28 days.

Results

First Response. The first response given by each *S* in each condition was scored as to whether it was the shocked or the nonshocked profile. Figure 2 shows the percentage of *S*s, in each condition, reporting the shocked profile

TABLE 4

ANALYSES OF VARIANCE FOR ESCAPE AND NO-ESCAPE CONDITIONS

Condition	Source	*df*	Mean square	*F*	*P*
Escape	18, 25, 50, 75 volts Within	3 39	32.20 9.81	3.28	<.05
No-escape	18, 25, 40, 50, 60 volts Within	4 59	16.97 7.42	2.28	<.10

TABLE 5

COMPARISON OF NUMBER OF SHOCKED FACES REPORTED IN ESCAPE AND NO-ESCAPE CONDITIONS

Volts	Escape Conditions	$\bar{X}$	$SD_{\overline{X}}$	*t*	*df*	*p*
18	Escape No-escape	9.62 7.00	.943 .966	1.94	13	$.10 > p > .05$
25	Escape No-escape	8.14 5.29	.702 .880	2.53	27	$.02 > p > .01$
50	Escape No-escape	8.33 7.13	.569 816	1.206	29	N.S.

with voltage level of the associated shock. This figure makes use of the data from Experiment I where 40 and 60 volts were presented under no-escape conditions. In the escape conditions, 24 out of 43 *S*s reported the nonshocked face, whereas in the no-escape conditions, 42 out of 64 *S*s reported the nonshocked face. Overall, 66 out of 107 *S*s reported the nonshocked face.

Total Response Scores. The total number of times, out of 14 possible, that *S* reported the shocked face was also taken as a score. The relationship between average number of shocked faces reported and associated voltage level is shown in Fig. 3. There is a difference between the various levels of shock at the .05 level for the escape group and at the .10 level for the no-escape group. This is shown in the analyses of variance summarized in Table 4. The same data were also analyzed by means of nonparametric median tests which confirmed the results of the analyses of variance.

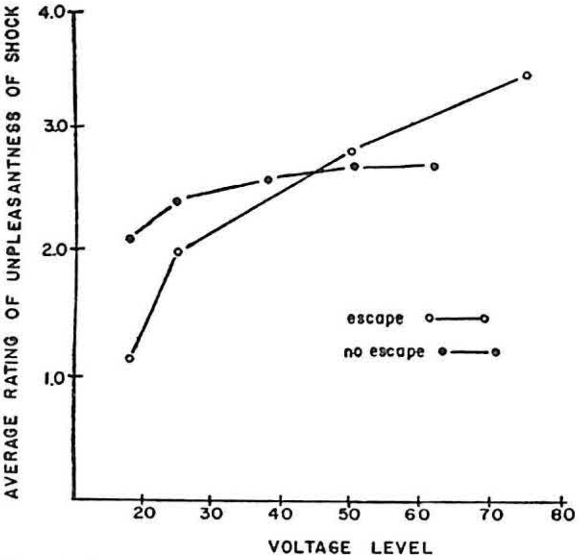

Fig. 4. Comparison of Subjective Rating of Shock with the Objective Voltage Level

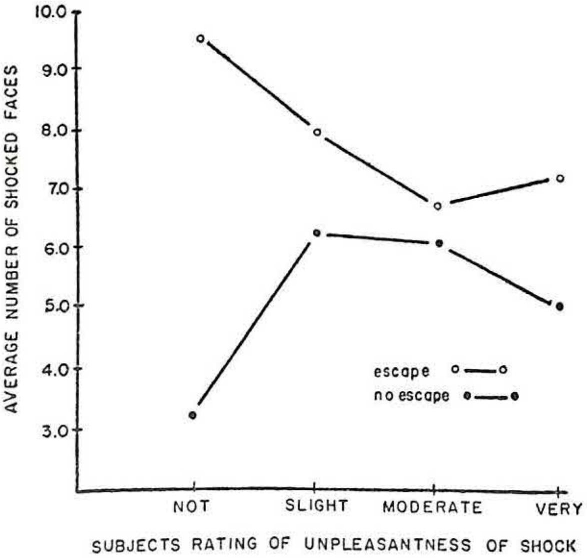

Fig. 5. Number of Shocked Faces Reported as a Function of Reported Unpleasantness of Shock

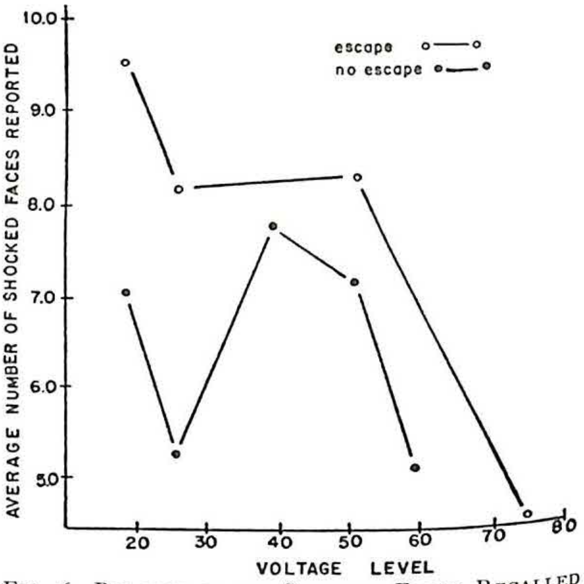

Fig. 6. Percentage of Shocked Faces Recalled Most Vividly by *S*s in Each Escape Condition

Table 5 summarizes the means and standard error of the means of the number of shocked faces reported for each of the conditions involved in this experiment, as well as *t* tests between escape and no-escape conditions at each level of shock.

Most experimenters do not state the physical values of the shock presented to *S* but only the fact that the shock was rated at a certain unpleasantness level. Figure 5 accordingly disregards voltage levels and shows the relationship between average number of shocked faces reported and judged unpleasantness. There is an almost perfectly symmetrical relationship on either side of a horizontal line, between the two variables for the escape and no-escape groups. Figure 4 shows the relationship between average rank unpleasantness and voltage administered. There is a strong, positive correlation between perceived or judged unpleasantness and the physical values of the shock for the escape group but little or no correlation between these two indices of punishment for the no-escape group.

Recall Analyses. Immediately following the test phase of the experiment, the *S*s had been asked which face they recalled most vividly, which next, and so on. Figure 6 shows the relationship between proportion of *S*s recalling the shocked face more vividly than the nonshocked and voltage level used in association. Both the escape and no-escape groups show a significant tendency to recall the shocked face as more vivid than the nonshocked up to 60 volts, after which level the nonshocked face is recalled most vividly. As physical intensity of shock increased in training, more and more nonshocked faces are recalled as more vivid.

Delayed Testing Analyses. No significant trends were found, either as to reporting the shocked or the nonshocked profiles, with delay in testing. The escape group tended slightly to report more shocked profiles following delay in testing, whereas the no-escape group tended slightly to report more nonshocked profiles. A few *S*s in each group showed marked shifts in perceptual reports during the second week after the training. However, these dramatic shifts in reporting either the shocked or nonshocked profiles seem to reflect individual differences rather than kind of training.

DISCUSSION

A brief review of the major results is in order. We found that reporting of the nonshocked face increased as a function of increasing severity of shock. When *S*s could escape from the electric shock they were more likely to "emphasize" the shock-associated percept than when they could not escape from the shock. There was a strong, positive correlation between judged unpleasantness of the shock and the physical voltage administered for the escape conditions, but no significant correlation for the no-escape condition. The shock-associated profile was recalled as most vivid by both escape and no-escape groups, this effect decreasing with increasing amounts of shock. With delayed testing finally, there were dramatic shifts toward reporting more nonshocked faces approximately in the second week after training, this trend being largely a function of individual differences rather than training conditions.

Certain weaknesses in our studies should be recognized at the outset. It would of course have been highly desirable to repeat all of our studies in the form of one multifactorial design. Another problem arises from the fact that the escape group reduced the duration of shock by their escape, so that the duration/intensity relationship was not the same for escape and no-escape conditions. Finally, and perhaps more serious, the question can reasonably be asked: "Is this perceptual learning?" It is clear that our design and data do not guarantee that we were changing "perception" rather than verbal habits. We can only point out that the consistency of our findings suggests to us that something more than mere verbal habits was altered, and there are enough indirect hints to suggest that "perception" of the tactual plaques changed as a function of the punishment association.

Our data agree with Smith and Hochberg (11), Dulany (4), Pustell (7), and Reece (10) in that severe punishment, i.e., electric shock, disrupts or interferes with perceptual organization, especially when the *S* cannot escape from the punishing event. We have gone further and have shown that with mild shock and escape conditions an "alerting" effect is obtained in that the punished percept is emphasized in a tactual figure-ground organization. Indeed, at all levels of shock, escape seems to make the *S* more alert to the associated percept than does no-escape. One is much more likely to obtain an "autistic" organization (5) when the *S* cannot overtly escape from the shock than when he can.

Our data also shed light on the problem of whether to use subjective evaluations of unpleasantness of shock or to use physical measures of shock. Subjective evaluations and physical measures seem to be interchangeable, functionally, under escape conditions. There is such a strong, positive correlation between the two in the escape condition that they must be reflecting the same psychological phenomenon. In the no-escape condition, however, there is virtually no correlation between the two and they are not functionally interchangeable. It is highly possible that in the no-escape condition, *S*s are rating the *total* unpleasantness of the situation, including the unpleasantness of being strapped into the elastic glove. There are other complications, too, in that adaptation to the shock is more probable in the no-escape condition than in the escape condition.

Our use of the concept "repression" to make predictions concerning delayed testing results warrants some comment. Dollard and Miller

(3) have pointed out that perceptual cues can acquire reinforcing properties through past association with punishment and produce behavior similar to that elicited by a primary negative stimulus. In our study, the percept associated with shock should thus acquire secondary aversive properties, which might appear upon delayed testing. Rapaport (8) has emphasized that anxiety associated with percepts increases after an unpleasant experience and that these percepts become "repressed," the repression increasing with the passage of time. The term "repression" is used here in a purely descriptive sense. Our results indicate that "repression" does not immediately develop following the shock but sets in about two weeks afterwards. The mechanisms underlying repression of the shock-associated percept remain to be clarified.

Summary

Three experiments are reported in which tactual profiles of faces were associated with electric shock of either 18, 25, 40, 50, 60, or 75 volts administered to the unpreferred hand, not used in tracing the plaque. In Experiments I and III the shock was temporally contiguous with the traced profile, and in Experiment II it was given three seconds afterwards. In some conditions, the subjects were allowed to escape from shock and in others they were not. Immediately following the training phase, in which the shock was associated with one tactual profile and no-shock with the alternative profile, the subject was blindfolded and traced an "ambiguous" line (which was the identical contour for both the shocked and the nonshocked profiles), reporting which profile was being traced. Subjects also rated the unpleasantness of the electric shock and reported which face they remembered most vividly.

The data were analyzed with the following results:

1. As intensity of shock increases, there is more and more reporting of the nonshocked profile in the test situation.

2. Escape conditions lead to more reporting of the shocked profiles than do the no-escape conditions, i.e., escape is more likely to lead to emphasis of the shocked profiles in the tactual figure-ground organization than is no-escape.

3. For nearly all levels of shock, except 60 and 75 volts, the shocked profile is recalled most vividly.

4. There is a strong, positive correlation between subjective evaluation of the unpleasantness of the shock and the physical value of the shock in the escape condition, but there is little or no correlation between these two variables in the no-escape condition.

5. When testing with the ambiguous line is delayed (up to three weeks), there is more and more reporting of the nonshocked profile. This effect was not significant overall but was significant for some individuals.

REFERENCES

1. Ayllon, T., & Sommer, R. Autism, emphasis, and figure-ground perception. *J. Psychol.*, 1956, **11**, 163–176.
2. Bindra, D., & Cameron, L. Changes in experimentally produced anxiety with the passage of time: Incubation effect. *J. exp. Psychol.*, 1953, **45**, 197–203.
3. Dollard, J., & Miller, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
4. Dulany, D. G., Jr. Avoidance learning of perceptual defense and vigilance. Unpublished doctoral dissertation, Univer. of Michigan, 1954.
5. Murphy, G. Affect and perceptual learning. *Psychol. Rev.*, 1956, **63**, 1–15.
6. Postman, L., & Brown, D. R. Perceptual consequences of success and failure. *J. abnorm. soc. Psychol.*, 1952, **47**, 213–221.
7. Pustell, T. Cue and drive aspects of anxiety in relation to perceptual vigilance and defense. Unpublished doctoral dissertation, Univer. of Michigan, 1954.
8. Rapaport, D. *Emotions and memory*. Baltimore: Williams & Wilkins, 1942.
9. Razran, G. The dominance—contiguity theory of the acquisition of classical conditioning. *Psychol. Bull.*, 1957, **54**, 1–46.
10. Reece, M. M. The effect of shock on recognition thresholds. *J. abnorm. soc. Psychol.*, 1954, **49**, 165–172.
11. Smith, D. E., & Hochberg, J. E. The effect of "punishment" (electric shock) on figure-ground perception. *J. Psychol.*, 1954, **38**, 83–87.
12. Tolman, E. C. *Purposive behavior in animals and men*. New York: Century, 1932.

Received June 28, 1957.

INTRALIST SIMILARITY AND VERBAL ROTE LEARNING PERFORMANCE OF SCHIZOPHRENIC AND CORTICALLY DAMAGED PATIENTS[1]

ROBERT C. CARSON[2]

Northwestern University

THE problems with which this study is concerned are rooted in controversies regarding the performance of schizophrenic and brain-damaged individuals on certain intellectual tasks purporting to measure "abstracting" ability. On such tasks, organics and schizophrenics perform differently from normals and on an inferior level. Clinicians generally agree that organics do so because they are affected with what Goldstein (6, 8) has termed a loss of the abstract attitude; they cannot perceive similar or common elements, and every object becomes a concrete, unique entity divorced from a class. Less agreement exists as to the nature of the difference between schizophrenic and normal performance on this type of task. It is probable, however, that a majority of clinicians favor the view expressed by Cameron (1, 2), who conceives the schizophrenic as a person whose classes are too broad and inclusive owing to a functional thought disturbance.

One implication of these views is that organic patients suffer from an inability to make a common response to stimuli which are similar, while schizophrenic patients suffer from the opposite deficiency of being unable to make differential responses to stimuli which are different. Some recent investigators have reasoned from this that the concept of stimulus generalization (which refers to the process whereby a response regularly elicited, through training, by one stimulus tends to be elicited without prior experience by other, similar stimuli) might be useful in providing a more precise and potentially productive explanation of this type of pathological behavior than has heretofore been achieved. If this argument is sound, it implies that organics should exhibit less stimulus generalization (SG) than normals, while schizophrenics should exhibit more SG than normals. Garmezy (4) and Mednick (14) have both presented results indicating that in a simple S-R situation schizophrenics do in fact exhibit more SG than normals, although in the latter study the findings were not definitive. In a second portion of his study, Mednick found impressive confirmation of the hypothesis that cortically damaged organics exhibit less SG than do normals.

The results suggest, then, that distortion in the gradient of SG may be one determinant of the intellectual disturbances observed in schizophrenia and organic brain damage. It is the purpose of the present study to examine the performance of schizophrenics and cortically damaged organics in a verbal rote learning task, attempting to predict certain aspects of their performance by means of these notions about SG.

Modern interpretations of the nature of verbal rote learning are generally based on the formulations of Gibson (5). Typically, the list of items consists of a set of stimuli and a set of responses to be correctly associated with these stimuli. To the extent that the stimuli are similar (formally or in meaning), the response to one stimulus tends to generalize to (be evoked by) other stimuli in the list in proportion to the degree of similiarity between them. This generalization results in the occasional blocking of correct responses by interference from other responses with a consequent retardation of performance. Obviously, the extent to which generalization tendencies of this type develop (and hence, other things being equal, the difficulty of the list) is a function of intralist stimulus similarity. This interpretation of the verbal learning process has held up well experimentally and many observations are integrated by means of Gibson's conceptual framework (19).

---

[1] This report is based on a dissertation submitted to the Graduate School of Northwestern University in partial fulfillment of the requirements for the degree of doctor of philosophy. Thanks are due to Janet A. Taylor, dissertation adviser, for her generous help and encouragement during the investigation.

[2] Now at Department of Psychiatry, University of Chicago School of Medicine.

From what has been said regarding SG in schizophrenic and cortically damaged patients, it follows that these two groups should, on a verbal rote learning task, evidence differential rates of increase in difficulty with increasing intralist similarity: This rate of increase should be relatively high for schizophrenics, who show much SG; and it should be relatively low for cortically damaged organics, who show little SG. Specifically, the hypotheses of this study are as follows:

(*a*) As intralist similarity increases, schizophrenics exhibit a more pronounced rate of increase in difficulty than a normal control group.

(*b*) As intralist similarity increases, cortically damaged organics exhibit a less pronounced rate of increase in difficulty than a normal control group.

It should be noted that nothing is said here about relative levels of performance. Only differences in the *slope* of the performance curve when intralist similarity is varied, or an interaction between lists varying in degree of intralist similarity and neuropsychiatric condition (organic–schizophrenic–normal) of the *S*s, are predicted.

METHOD

Subjects

Three groups of male *S*s, distinguished by neuropsychiatric condition, were employed. A schizophrenic (Sc) group, a cortically damaged organic (O) group, and a neuropsychiatrically normal (N) group, each containing 45 *S*s, were selected on the basis of diagnostic and other criteria described below from the current patient populations of Veterans Administration hospitals in the Chicago area. Suspected mental defectives, persons having serious uncorrected visual defects, and persons having less than eight years of formal education were excluded.

Schizophrenics. The Sc group was composed of patients having a well-established diagnosis of schizophrenic reaction. The diagnosis was considered well-established when there was no evidence in the case folder of any disagreement in arriving at it. An attempt was made to include only relatively early chronic cases. Selection was limited to patients within the 20–50-yr. age range whose first psychiatric hospital admission occurred not less than 6 mos. nor more than 10 yrs. previously. No patient was included who had had more than 35 electroconvulsive shock or insulin coma treatments, or any assaultive somatic therapy within 3 mos. prior to participation in the study. In addition, no patient was included in whose case folder there was any suggestion of the possible presence of organic brain pathology or chronic alcoholism. For the entire Sc group, the mean age was 32.84 yrs.; the mean amount of formal education completed was 12.02 yrs., and the mean WAIS vocabulary score was 41.78.

Organics. Only patients with a well-established diagnosis of cortical brain damage were included in the O group. This diagnosis was considered well-established in cases where there had been brain surgery, or where definite neurological findings were substantiated by at least one of the common ancillary methods of neurological diagnosis (e.g., electroencephalography). It was originally hoped that the 20–50-yr. age criterion could be observed with reference to this group, but this proved impossible because of sheer lack of availability. The age range of the O group was 24–77 yrs., with a mean of 55.18 yrs. Any patient whose case folder had in it material which suggested that he was currently a serious psychiatric problem or that he had ever in the past been treated for a problem of this kind was excluded. Also excluded were patients exhibiting aphasic symptomatology. The O group had a mean WAIS vocabulary score of 36.09, and had completed, on the average, 9.64 yrs. of formal education.

Normals. The N group was composed of general medical and surgical patients who had received neuropsychiatric clearance from the physicians having responsibility for their cases. No patient who was known to have been treated for a psychiatric illness or for an intracranial neurological condition was included. Also excluded were known chronic alcoholics. The age range of the N group was 20–50 yrs., the mean being 35.69 yrs. The means for WAIS vocabulary score and amount of formal education were 36.71 and 10.62 yrs. respectively.

Materials

Three serial word lists of nine adjectives each, shown in Table 1, were constructed, using only words which could be assumed to be familiar to the *S*s. The average familiarity values of the three lists, as measured by Thorndike-Lorge (18) word counts, were approximately equal. The three lists differed in degree of intralist similarity of meaning. Roget's *Thesaurus* (15) was used as a reference for deciding similarity of meaning.

The low similarity list (I) consisted of nine adjectives which appeared to be quite dissimilar to each other in meaning, sound, and formal characteristics. The medium similarity list (II) consisted of four groups (three with two members and one with three members) of adjectives, the members of each group being similar in meaning to each other but dissimilar in meaning to the members of the other groups. The high similarity list (III) consisted of two groups (one with five members and one with four) of adjectives, each containing words selected on the basis of their similarity of meaning. Intralist similarity in dimensions other than meaning was minimized as much as possible in the construction of both this and the medium similarity list. A practice list, also reproduced in Table 1, consisting of seven dissimilar three-letter nouns was also employed. The arrangement of items within each of the lists was unsystematic.

All lists were printed in India ink on endless white tapes in upper-case letters approximately ½-in. high. They were presented by means of a Hull-type memory drum. To indicate the beginning of the list, a starter item (000) was printed at the top of each. There was a

TABLE 1

EXPERIMENTAL MATERIALS

Practice list	Experimental lists		
	I	II	III
Cow	Long	Gentle	Quick
Leg	Warm	Small	Merry
Map	Brave	Tiny	Joyful
Key	Tight	Moist	Fast
Bar	Holy	Rare	Gay
Son	Sour	Mild	Rapid
Hit	Extra	Scarce	Swift
	Formal	Wet	Happy
	New	Little	Speedy

2-sec. rate of exposure of items with an 8-sec. intertrial interval (12 sec. in the case of the practice list).

Procedure

Each of the three main groups (Sc, O, and N) were subdivided into three subgroups, each containing 15 *S*s. Each of the three subgroups within a main group learned a different experimental list of serial adjectives, described above. Certain performance criteria, to be outlined later, were established to determine the eligibility of *S*s to serve in the experiment. Subjects continued to be tested until 15 usable *S*s were obtained for each of the subgroups. Formation of subgroups was accomplished by alternate assignment of *S*s as they appeared to be tested. An exception to this general rule was made where there was clear evidence that this procedure would result in excessive variation among subgroups in practice list performance. Assignment of *S*s was then manipulated to obtain subgroups matched on practice list performance. There were, then, nine subgroups of 15 *S*s each, each subgroup representing a different combination of neuropsychiatric condition and experimental list learned. Age, education, and vocabulary score differences among subgroups within any main group were negligible.

After preliminary instructions, each *S* was given 10 learning trials on the practice list in order (*a*) to familiarize *S*s with and afford them some practice in verbal rote learning of this type, (*b*) to aid in further selecting *S*s by providing a basis for excluding those who were not likely, for whatever reason, to have any success in performing with the experimental lists, and (*c*) to afford a method of equating the three subgroups within each main group. If *S* failed to respond on the second practice list trial, *E* said, "Go ahead." Any *S* requiring help beyond this was excluded from the study. Only those *S*s who were unable on the 10th trial or before to produce in their correct sequential placement four of the seven words of the practice list were excluded on the basis of slowness of learning. Inadequate practice list performance resulted in the exclusion of 5 Normals, 13 Schizophrenics, and 10 Organics. *S*s who succeeded in rendering two consecutive perfect recitations of the practice list before the 10th trial were given credit for their remaining trials and allowed to proceed to the next list.

Approximately 1 min. after completion of the practice list trials, each *S* meeting the selection criteria was given 20 trials on the experimental list appropriate for the subgroup of which he was a member. *S*s who succeeded in rendering two consecutive perfect recitations of the experimental list within the 20 trials were discontinued at that point and given credit for perfect performance on their remaining trials. Following performance on the experimental list, every *S* was given the Vocabulary subtest of the Wechsler Adult Intelligence Scale to provide a rough index of his intellectual level.

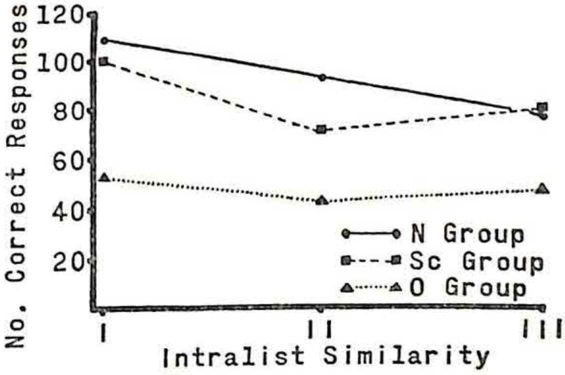

FIG. 1. MEAN NUMBER OF CORRECT ANTICIPATIONS AS A FUNCTION OF INTRALIST SIMILARITY

RESULTS

Practice list. The data of the experiment were organized in terms of the number of correct anticipations made by *S* in the allotted number of trials. The practice list data showed considerable variation in performance among the three main groups, but the performances of the subgroups within each main group were substantially equivalent. Appropriate analyses of variance revealed the following: (*a*) the practice list performance of the N group was, on the whole, significantly superior to that of the O group ($p < .001$) and of the Sc group ($p < .01$), and (*b*) as expected, there were no significant differences among the practice list performances of subgroups differentiated according to the experimental lists subsequently learned, thus permitting analysis of the effects of the experimental lists.

Experimental lists. Learning curves prepared separately for each of the nine subgroups indicated that the *relative* performance levels of subgroups remained fairly stable throughout the course of the 20 learning trials on the experimental lists. Accordingly, only the data for performance over the entire 20 trials were analyzed. The mean number of correct anticipation made on the appropriate experimental list was computed for each subgroup; these means are presented graphically in Fig. 1. Since practice and experimental list performance were found to be correlated ($r = .59$ for all 135 *S*s), analysis of covariance was applied where appropriate in the statistical evaluation of the experimental list data. Tests for homogeneity of variance and for homogeneity of regression of experimental on practice list scores failed to show significant heterogeneity of either type among the subgroups.

Organics vs. normals. It was hypothesized that, as intralist similarity increases, the O group exhibits a less pronounced rate of increase in difficulty than the N control group. The hypothesis seems to be supported by the results shown in Fig. 1. Contrary to expectations, however, an analysis of covariance performed on the O and N group data showed the interaction term for neuropsychiatric (NP) conditions × lists to be short of statistical significance, *F* having a probability of approximately .15. The list effects and the NP condition effects, taken independently, were both highly significant ($p < .005$ in both cases). In view of the use of analysis of covariance procedures, the significant difference between the O and N condition is of special interest because it represents a difference between the O and N groups in performance on the experimental lists *after* their scores on these lists have been adjusted to equality on the basis of practice list performance. This means in effect that the difference in difficulty between the practice and experimental lists was generally greater for the O group than for the N group, even though the coefficient of regression of experimental on practice list scores was not significantly different for the two groups.

Failure to find a significant NP conditions × lists interaction represents, of course, a failure to obtain support for the hypothesis. Nevertheless, it must be pointed out that the interaction term, while not reaching a conventionally acceptable significance level, does approach it rather closely. The obtaining of a significant interaction term in a statistical test, particularly where, as in the present case, the test is made on relatively small samples from highly variable populations, is notoriously difficult. A supplementary method of testing the hypothesis was therefore employed.

The experimental list data were grouped separately for the N and O conditions, and a simple analysis of variance was performed to test list effects in each condition. Summaries of these analyses appear in Table 2. For the N group, list effects are highly significant ($p < .01$), while for the O group, the list effects do not approach a statistically significant level, *F* being less than 1.00. Thus, it would appear that for the O group, difficulty does not vary significantly with changes in intralist similarity, whereas for the N group there is a highly significant increase in difficulty with increasing intralist similarity.

While these analyses appear to confirm the hypothesis of less generalization in the O than in the N group, it could be argued that the lack of difference among the O subgroups is due to their general and pronounced inferiority; *S*s learning the medium and high similarity lists could not, according to this interpretation, be much worse than those in low similarity group. In order to test this possibility, the five *S*s having the *lowest* experimental list scores in each of the three N subgroups, and the five *S*s having the *highest* experimental list scores in each of the three O subgroups were selected for special study. Practice list scores for these *S*s were subjected to analyses of variance, in which no significant differences were found between subgroups within either the high O or low N groups. The mean number of correct anticipations on the appropriate experimental list was computed for each of the six selected subgroups, and inspection revealed that, although the high Os and low Ns were roughly equal with respect to general level of performance, the means of the Ns decreased in regular fashion as intralist similarity increased, while those of the Os did not. This impression was confirmed by analyses of variance performed on these data: for the low Ns, list effects were highly significant ($p < .001$), whereas for the high Os, list effects were not significant, *F* again being less than 1.00. Thus, the general O and N group findings were duplicated in this miniature analysis employing selected *S*s. One may conclude, then, that the hypothesis is essentially supported. As meaningful intralist similarity increases, cortically damaged organics exhibit a less pronounced rate of increase in difficulty than do neuropsychiatrically normal *S*s. Indeed, the results suggest that the organics exhibit *no* significant increase in difficulty with increasing intralist similarity. This effect, moreover, appears to be independent of the general level of performance.

TABLE 2

SUMMARIES OF ANALYSES OF VARIANCE OF EXPERIMENTAL LIST SCORES FOR NORMAL, ORGANIC, AND SCHIZOPHRENIC GROUPS

Group	Source of variation	Degrees of freedom	Mean square	*F*	*p*
Normals	Between subgroups (lists)	2	3811.82	6.05	$<.01$
	Within subgroups	42	630.48		
	Total	44			
Organics	Between subgroups (lists)	2	431.62	.54	—
	Within subgroups	42	805.81		
	Total	44			
Schizophrenics	Between subgroups (lists)	2	3323.27	3.00	$>.05$
	Within subgroups	42	1108.08		
	Total	44			

Schizophrenics vs. normals. It is apparent from Fig. 1 that the performance of the Sc group on the experimental lists unexpectedly approximates that of the N group. Only on List II did the difference appear very great, and even here it was not statistically significant. An analysis of variance performed on the experimental list data for the N and Sc group showed no significant NP condition effects ($p > .10$). The N and Sc groups therefore cannot be regarded as differing with respect to level of performance on the experimental lists.

The hypothesis stated that as intralist similarity increases, the Sc group exhibits a more pronounced rate of increase in difficulty than N control *S*s. An analysis of covariance showed the *F* for the NP conditions × lists interaction to be far below an acceptable level of significance ($p > .20$). Moreover, it is clear from inspection of the data and from the supplementary analysis described below that any interaction occurring here was in a direction opposite to that predicted. As expected, the *F* for NP conditions was not significant, whereas that for lists was highly significant ($p < .001$).

It will be recalled that when the data for the N group alone were subjected to analysis of variance, the list effects were found to be highly significant (Table 2). The implication of the hypothesis is that list effects should be greater in the Sc than in the N group. A simple analysis of variance, summarized in Table 2, shows no significant list effects ($p > .05$). Consequently, the hypothesis seems untenable. The Sc group did not show a greater than normal increase in difficulty with increasing intralist similarity; like the O group, it showed none.

DISCUSSION

Organics. The O group results confirm the hypothesis with respect to that group, thus supporting the theoretical framework from which it was derived. The pathologically lowered gradient of SG, which may be a strong factor in limiting the performance of organics on many intellectual tasks (e.g., those concerned with concept formation), is actually conceived here as an aid to performance. Furthermore, if the views presented here are correct, cortically damaged *S*s should have a similar advantage on any complex task in which, because of SG, interference from incorrect competing responses is a major deterrent to successful performance. It is not suggested that they will perform better than normals, for SG is obviously not the only process altered in cortical injury. It may be hypothesized, however, that performance differences between normal and cortically damaged *S*s on a given task will become less as the opportunity is increased for SG-produced interference to occur.

It has been shown that the empirical gradient of SG is positively related to such factors as the habit strength of the original S-R association and to the drive level currently operative (9). Mednick (14) had the impression that his organic group was generally low in drive. On the assumption that amount of SG exhibited varies directly with drive level, he suggested that the lowered gradient of SG in cortically damaged *S*s may be due to a pathologically diminished level of drive. In the present study, *E* had no such impression concerning the drive level of the O group. Moreover, the analysis of the equated groups challenges a drive interpretation. Os and Ns who were approximately equal in general level of performance were selected for study. The findings could issue from differences between the N and O groups in learning, in drive, or both. If either of these factors were responsible, however, equating groups, as was done, would be expected to result in the disappearance of the effect in question. No such result was found.

In the absence of more reliable evidence, observations made here favor the hypothesis that the cortically injured individual is deficient in some kind of basic generalizing ability. This is, of course, essentially in keeping with Goldstein's original notions about the effects of brain injury, although the issues are probably more complex than he conceived them to be. In any case, the results suggest that the concept of SG may offer an explanatory principle of wider generality than Goldstein's abstract-concrete dichotomy.

As noted above, the statistical analysis indicated, unexpectedly, that the difference in difficulty between the practice and experimental lists was greater for the O than for the N group. Undoubtedly, the experimental lists, because of their length and composition, are on the whole more productive of interference than the practice list. But the results indicate that cortically damaged *S*s are less susceptible to such interference than are normal *S*s; accordingly, an interference hypothesis does not afford an adequate explanation. The explanation probably lies in the common observation that brain-damaged individuals often show a decrement in performance with sustained effort. Such an effect was probably operative in the O group as they moved from the practice through the experimental lists, thus accounting for their disproportionately poorer performance on the latter.

Schizophrenics. It was hypothesized that with increasing intralist similarity, the Sc group would show progressively poorer performance than the N group. Actually, the performance of the Sc group did not vary significantly with intralist similarity. The results obtained for the Sc group are more in accord with the views held by Vigotsky (20), Goldstein (7), and Kasanin and Hanfmann (11), according to which the primary psychological disturbance in schizophrenia is a concretization of mental processes or, in other words, a deficiency in the ability to abstract and conceptualize. These investigators believe that the schizophrenic, like the organic, is more or less incapable of appreciating similarity among different stimuli. The present results are consistent with such an hypothesis. However, the findings on SG in schizophrenia engender a dilemma: if the schizophrenic is indeed unable to appreciate similarity among different stimuli, he should exhibit less SG than the normal *S*; two independent investigations (4, 14) indicate that quite the opposite is true. Furthermore, the character of the observations which led to the application of the concretization hypothesis to schizophrenia is highly questionable, having been almost exclusively of a qualitative and uncontrolled type. Recent investigations of schizophrenic conceptual performance, in which more rigorous and sophisticated experimental procedures were employed, have failed to confirm the concretization notion (3, 12, 13). One is therefore extremely reluctant to conclude that the results reflect a concretization of mental processes or a lack of generalizing ability in the Sc group.

Nor do the findings seem accountable on the basis of any systematic knowledge concerning the gradient of SG. The high performance level of the Sc group all but rules out the possibility of a decrement either in habit strength or in drive. There remains only the rather uncomfortable position of having to suggest that the findings may be due to the influence of some uncontrolled artifact operating in connection with the Sc group. One possibility which comes immediately to mind is that of an uncontrolled medication effect. A considerable majority of the patients in the Sc group were under the influence of chlorpromazine at the time of testing. It is not inconceivable that such medication may depress the gradient of SG, and this would account for the ineffectiveness of the intralist similarity variable in the performance of the Sc group.

It has been noted previously that the general level of performance of the Sc group on the experimental lists was not significantly inferior to that of the N group. This finding runs counter to much experimental data showing schizophrenics almost without exception to be inferior to normals on intellectual tasks (10, 16). The explanation may lie, at least in part, in the fact that the Sc group was somewhat more intelligent and better educated than the N group. These differences may be a function of variation in the patient populations of the hospitals at which the *S*s were obtained.

On the practice list, the Scs conformed to expectations in performing at a significantly lower level than the Ns. This raises the question of why there should be a difference between the two groups on the practice list but no significant difference between them on the experimental lists. Two explanations seem plausible: (*a*) the practice list was more difficult for the Sc group than for the N group due, as some evidence (17) would suggest, to the former's slowness in adjusting to a novel situation; or (*b*) assuming a genuine superiority in the N group, the experimental lists were more difficult for the N than for the Sc group due to the N group's greater susceptibility to interference. There seems to be no reasonable basis for deciding between these alternative explanations.

Finally, a word is perhaps in order concerning the possibility of relevant but uncontrolled variables being operative in connection with the composition of the experimental and control groups. The most conspicuous problem is the discrepancy in mean age between the N and O groups, the O group being on the whole considerably older. While it is probable that the most significant changes accompanying advanced age are in the direction of increasing organic deterioration, there is the very remote possibility that the O group results are more a function of age than of organic pathology per se. Some further difficulties of a similar nature arise in connection with differences among groups in intelligence, education, and length of hospitalization. In general, Scs and Os were long-term, chronically hospitalized cases, whereas Ns tended to be acute cases only recently hospitalized. Again here, however, the possibility of influence from these factors on the main findings is considered remote.

SUMMARY

Experimental work growing out of earlier notions about conceptual ability in schizophrenic and brain-injured individuals has shown distortions of the gradient of stimulus generalization in these two pathological groups. Stated simply, it has been shown that schizophrenics exhibit more stimulus generalization than normals, while persons with cortical brain damage exhibit less stimulus generalization than normals. These findings may help to explain certain anomalies in the intellectual performance of schizophrenic and brain-injured individuals.

The findings on stimulus generalization in these pathological conditions, in view of current theoretical interpretations of the role of stimulus generalization in producing the relationship between intralist similarity and difficulty of verbal rote learning, led to the adoption of the following hypotheses. With increasing intralist similarity in the material learned, (*a*) schizophrenics exhibit a more pronounced rate of increase in difficulty than a normal control group, and (*b*) cortically damaged organics exhibit a less pronounced rate of increase in difficulty than a normal control group.

Three groups of hospitalized patients, diagnosed respectively as schizophrenic, cortically injured, and neuropsychiatrically normal, were each divided into three equated subgroups. Each subgroup within a main group learned one of three serial adjective lists which differed from each other in degree of intralist similarity of meaning. Thus, there were nine subgroups, each representing a different combination of neuropsychiatric condition and degree of intralist similarity in the experimental list learned.

Analysis of results supported Hypothesis (*b*). Hypothesis (*a*) was unconfirmed; the results for the schizophrenic group were opposite to those predicted by the hypothesis. Like the organics, they showed a *less* than normal increase in difficulty with increasing intralist similarity. However, various considerations prevent the drawing of any conclusions from this finding.

REFERENCES

1. CAMERON, N. Schizophrenic thinking in a problem-solving situation. *J. ment. Sci.*, 1939, **85**, 1012–1035.
2. CAMERON, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. II. New York: Ronald, 1944. Pp. 861–921.
3. CHAPMAN, L. J. The role of type of distractor in the "concrete" conceptual performance of schizophrenics. *J. Pers.*, 1956, **25**, 130–141.
4. GARMEZY, N. Stimulus differentiation by schizophrenic and normal subjects under conditions of reward and punishment. *J. Pers.*, 1952, **20**, 253–276.
5. GIBSON, ELEANOR J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
6. GOLDSTEIN, K. *The organism*. New York: American Book, 1939.
7. GOLDSTEIN, K. The significance of psychological research in schizophrenia. *J. nerv. ment. Dis.*, 1943, **97**, 261–279.
8. GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior: An experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
9. HULL, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.

10. HUNT, J. McV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. II. New York: Ronald 1944. Pp. 971–1032.
11. KASANIN, J., & HANFMANN, EUGENIA. Disturbances in concept formation in schizophrenia. *Arch. Neurol. Psychiat.*, 1938, **40**, 1276–1282.
12. MCGAUGHRAN, L. S., & MORAN, L. J. "Conceptual level" vs. "conceptual area" analysis of object-sorting behavior of schizophrenic and non-psychiatric groups. *J. abnorm. soc. Psychol.*, 1956, **52**, 43–50.
13. MCGAUGHRAN, L. S., & MORAN, L. J. Differences between schizophrenic and brain-damaged groups in conceptual aspects of object sorting. *J. abnorm. soc. Psychol.*, 1957, **54**, 44–49.
14. MEDNICK, S. A. Distortions in the gradient of stimulus generalization related to cortical brain damage and schizophrenia. *J. abnorm. soc. Psychol.*, 1955, **51**, 536–542.
15. ROGET, P. M. *Thesaurus*. New York: Grossett & Dunlap, 1947.
16. SHAKOW, D. The nature of deterioration in schizophrenic conditions. *Nerv. ment. Dis. Monogr.*, 1946, No. 70.
17. SHAKOW, D., & HUSTON, P. E. Studies of motor function in schizophrenia: I. Speed of tapping. *J. Gen. Psychol.*, 1936, **15**, 63–108.
18. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Columbia Univer., 1944.
19. UNDERWOOD, B. J. *Experimental psychology*. New York: Appleton-Century, 1949.
20. VIGOTSKY, L. Thought in schizophrenia. *Arch. Neurol. Psychiat.*, 1934, **31**, 1063–1077.

Received July 1, 1957.

THE INTELLIGENCE TEST PERFORMANCE OF MAORI CHILDREN: A CROSS-CULTURAL STUDY

RICHARD H. WALTERS[1]

University of Toronto

EDUCATORS have come to rely heavily on the measurement of ability by psychological tests for the prediction of academic success. In a community like New Zealand where the Maori form a considerable minority group of varying degrees of acculturation and where, for the most part, Polynesians share the same educational opportunities and advantages as Europeans, the routine testing of mixed groups may bring about misleading results. The investigation reported in this paper was carried out to assess the performance of Maori pupils on a variety of tests, verbal and nonverbal, of a kind customarily used for the assessment of intelligence among Euro-American populations.

THE TESTING PROGRAM

During the winter months of 1953, testing was carried out in Northern New Zealand with schoolchildren between the ages of 11.0 and 15.11. Two test batteries were used: (*a*) the Science Research Associates' shortened form of Thurstone's Tests of Primary Mental Abilities (PMA); and (*b*) a "Non-Language Test" (NLT) comprised of items from existing test batteries with modifications designed to reduce the effects of differing cultural experiences.

Maori testees were taken from three general sources: (*a*) from schools in the vicinity of Auckland under the supervision of the local education authority, with the addition of a small number of cases from the town of Whangerei (city group); (*b*) from semi-rural areas bordering on Auckland in which the Maori works as a laborer in market gardens and lives under extremely adverse social conditions (market-garden group); (*c*) from country schools under the supervision of the Department of Maori Schools (country group). Within the country group was included a small proportion of pupils from boarding schools in and near Auckland, since by far the greater number of these Maori pupils come to the boarding schools around the age of 13 or 14 from Maori primary schools in predominantly Maori areas and so are lost to the Maori District High Schools.

The control group consisted of children of European descent taken almost exclusively from the public schools of Auckland, the few exceptions being pupils of European descent who happened to be attending Maori schools for geographical reasons and a few pupils from Whangerei High School who were tested at the same time as the Whangerei Maori group. The selection of a group of pupils primarily from city schools to serve as a control group may have put the Maoris at a disadvantage, since there is a tendency for city pupils in New Zealand to score higher on intelligence tests than do country pupils (2). The selection of such a group was, however, necessary on account of limitations of time and finance.

The cooperation of school principals was enlisted for the administration of the tests. In addition, principals or classroom teachers were asked to rate each Maori pupil on each of five three-point scales designed to give a rough measure of variables that might be related to test scores: school attendance, continuity of education, conversational verbal fluency, degree of acculturation, and educational progress.

A total of over 900 Maori and 600 control children were tested. There is some doubt concerning the representativeness of the sample of 15-year-olds, since in New Zealand a pupil may leave school as soon as he has attained his fifteenth birthday. Throughout the 11-, 12-, 13-, and 14-year-old groups, however, adequate samples were obtained both of Maori and of control children, and results from one age group to another were remarkably consistent. Consequently, this paper is based primarily on the results for the 13-year-old group only.

Description of the Tests

As is well known, the PMA test consists of six subtests: Verbal Meaning, Space, Reasoning, Number and Word Fluency, corresponding to Thurstone's Factors V, S, R, N, and W. The PMA test was administered in a standard manner. However, because of the difficulty some Maori children appeared to find in the use of the answer pad, the form of test was modified to allow answers to be written beside the test item instead of being recorded on an answer pad. This modified form was used with pupils who did not reside in urban areas, but the standard answer pad form was used with pupils in the city schools who are more accustomed to testing procedures. An unpublished study by Leone Smith of Auckland University College has shown that the use of the modified PMA test booklet does not significantly change scores of Maori pupils, though it does significantly raise the scores of children of European descent. Since almost all the control group was tested with the standard answer pad form, it is unlikely that results have been seriously affected. The use of the modified form in country and semi-rural areas undoubtedly saved many spoiled test records.

---

[1] This research was carried out while the author was on the faculty of Auckland University College, New Zealand.

The non-language battery consisted of six subtests: a Substitutions test which was essentially the Digit Symbol Test of the Wechsler Intelligence Scale for Children, Picture Completion and Picture Arrangement tests based primarily on the Wechsler-Bellevue and WISC items, a Block Counting test consisting of items taken from the Army General Classification Test, a Differences test taken from the SRA Non-Verbal Intelligence Test, and a Series test based on items from the Differential Aptitude Tests. All subtests were set up so that they could be administered as a group intelligence test with time limits which were set after some preliminary experimentation.[2]

The Rating Scales

The five rating scales were each defined by three points in such a way that a school principal or classroom teacher might be able very quickly to give a reasonably accurate rating on the variables under consideration.

1. School Attendance:

   Rating 1. Regular attendance.

   Rating 2. Occasional absences; not excessive, but sufficient to impose some handicap.

   Rating 3. Attendance very irregular.

2. Continuity of Education:

   Rating 1. Attendance at one school only during school career. Rate 1 also if the only change has been from primary to postprimary school.

   Rating 2. Attendance at two, but not more, schools either at primary or postprimary level, or at both.

   Rating 3. Attendance of three or more schools either at primary or at postprimary stage.

3. Verbal Fluency (Conversational):

   Rating 1. Very fluent; good self-expression.

   Rating 2. Has reasonable command and understanding of English.

   Rating 3. Very little command of English.

4. Acculturation:

   Rating 1. Home conditions approximating to average European home. Rate 1 whether or not Maori is spoken at home provided good English is also spoken by parents.

   Rating 2. A good home with reasonable amount of European cultural influence. A good deal of Maori spoken; English of parents of mediocre quality.

   Rating 3. A poor home with little or no European influence. Very little, if any, English spoken. If conditions are extremely poor, rate 3 even if parents speak reasonably good English.

5. Educational Progress:

   Rating 1. Better than average.

   Rating 2. Average.

   Rating 3. Below average.

The use of broadly defined categories including evaluative terms was decided upon after consultation with teachers who had had experience of teaching in Maori schools. It was felt that such scales could be more easily, and would be more readily, used by teachers than ones which were more detailed and more objectively based. Since the scales were used by a considerable number of teachers, and the number of pupils who were rated is very large, an adequate rough grouping of pupils on the five variables has almost certainly been achieved.

RESULTS

Table 1 gives the mean scores of the control and the three Maori groups on the subtests, and also the total test result, for both batteries (13-year-old group only). A test of homogeneity of variance, using Bartlett's procedure, was carried out for the three Maori groups; the control group was then added and a further test of homogeneity was made. Whether or not heterogeneity was found, an analysis of

---

[2] Copies of the non-language test and of instructions given to administrators have been lodged with the American Documentation Institute. Order Document No. 5558 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $2.00 for microfilm or $3.75 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1

INTELLIGENCE TEST SCORES OF 13-YEAR-OLD MAORI AND CONTROL GROUPS

Primary Mental Abilities

Subtest	Control (N = 116)		Country Maori (N = 131)		City Maori (N = 48)		Market-Garden Maori (N = 36)	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Verbal Meaning	17.66	8.22	15.08	7.24	13.27	7.93	7.51	4.50
Space	19.58	8.79	11.05	9.97	14.46	11.56	9.11	7.75
Reasoning	11.41	5.88	8.50	5.00	7.31	4.62	6.64	4.33
Number	14.55	7.55	18.88	8.02	14.27	10.68	11.83	6.67
Word Fluency	35.24	11.23	36.74	11.97	36.10	13.22	29.83	11.20
IQ	102.69	16.74	101.25	15.62	96.65	14.97	87.22	13.05

Non-Language Test

Subtest	Control (N = 126)		Country Maori (N = 132)		City Maori (N = 49)		Market-Garden Maori (N = 37)	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Substitutions	14.91	3.30	14.83	5.16	14.94	3.12	13.30	5.92
Pict. Comp.	10.59	2.91	9.39	2.38	9.39	2.41	7.86	2.66
Pict. Arr.	12.36	3.47	9.29	3.55	9.57	3.30	7.73	3.83
Block Counting	12.75	3.57	10.80	3.00	10.74	2.77	9.49	3.28
Differences	14.22	2.70	12.91	3.10	12.90	4.12	11.62	3.00
Series	9.99	4.36	7.23	3.95	8.14	3.83	6.47	3.81
Total	74.82	13.58	64.45	11.78	65.67	13.03	56.49	12.19

variance was made both for the three Maori groups alone and for the control and three Maori groups together. Wherever the value of F was significant, tests of the significance of differences between mean scores were carried out, using the within variance as the best estimate of population variance when the variance of the samples was homogeneous and the critical ratio formula in cases where heterogeneity of variance had been found. Since the primary concern is to compare the Maori groups with each other, and to compare each of them individually with the control group, only the results of the t tests are reported here (Tables 2 and 3).

The results for the 13-year-old group are fairly typical of what was found at each age level. On the PMA, the country Maori does almost as well as the control, as far as IQ is concerned. On the other hand, the control group does significantly better than the city or the market-garden Maori, which appears to be definitely an intellectually inferior group.

The pattern of scores within the PMA is of especial interest. In general, the control group tends to do better than all three of the Maori groups on V, R, and S; on the other hand, the country Maori performs decidedly better than the control group on N, and does equally well on W. In fact, the city and country Maori groups showed a greater increase in W over the ages studied than did the control group, so that, after gaining a somewhat lower score than the control group on this factor at the 11-year level, by the 15-year level they obtain a score which is significantly higher than that of the control group.

The picture is quite different for the NLT on which the Maori might have been expected to do relatively better than on the PMA. At all age levels, the control group gained a significantly higher total score than any of the Maori groups and showed definite superiority on all subtests except Substitutions. The differences between the country and city Maori largely disappear; the groups do about equally poorly compared to the controls. The market-garden group, however, still remains a much inferior group in test performance.

The tests of homogeneity of variance provided important additional data. On the PMA, there was no case for the 11- through 14-year-old groups in which the addition of the control group resulted in heterogeneity when the variances of the Maori group had been

TABLE 2

SIGNIFICANCE OF MEAN GROUP DIFFERENCES BETWEEN INTELLIGENCE TEST SCORES OF COUNTRY, CITY, AND MARKET-GARDEN MAORI GROUPS

Primary Mental Abilities

Subtest	Country vs. City (*df* = 177)		Country vs. Market-Garden (*df* = 165)		City vs. Market-Garden (*df* = 82)	
	Diff.	*t*	Diff.	*t*	Diff.	*t*
Verbal Meaning	1.81	1.38	7.27	7.41**	5.46	3.99**
Space	−3.41	1.81	1.94	1.24	5.35	2.54*
Reasoning	1.19	1.46	1.86	2.05*	0.67	0.81
Number	4.61	2.72**	7.05	5.37**	2.44	1.28
Word Fluency	0.64	0.31	6.91	3.02**	6.27	2.34*
IQ	4.60	1.81	14.03	4.94**	9.43	2.83**

Non-Language Test

Subtest	Country vs. City (*df* = 179)		Country vs. Market-Garden (*df* = 167)		City vs. Market-Garden (*df* = 84)	
	Diff.	*t*	Diff.	*t*	Diff.	*t*
Pict. Comp.	−.01	0.01	1.53	3.19**	1.52	2.71**
Pict. Arr.	−.28	0.46	1.56	2.26*	1.84	2.28*
Total	−1.23	0.59	7.96	3.47**	9.19	3.42**

Note.—In the preliminary analyses of variance, the *F* ratios for Substitutions, Block Counting, Differences, and Series were not significant; consequently, *t* ratios were not calculated.

* Significant at the .05 level.

** Significant at the .10 level.

TABLE 3

SIGNIFICANCE OF MEAN GROUP DIFFERENCES BETWEEN INTELLIGENCE TEST SCORES OF CONTROL AND MAORI GROUPS

Primary Mental Abilities

Subtest	Control vs. Country Maori (*df* = 245)		Control vs. City Maori (*df* = 162)		Control vs. Market-Garden Maori (*df* = 150)	
	Diff.	*t*	Diff.	*t*	Diff.	*t*
Verbal Meaning	2.58	2.60**	4.39	3.19**	9.85	9.19**
Space	8.53	4.64**	5.12	1.14	10.47	8.53**
Reasoning	2.91	4.37**	4.10	4.58**	4.77	4.79**
Number	−4.33	4.58**	0.28	0.17	2.72	2.07*
Word Fluency	−1.50	0.99	−0.86	0.44	5.41	2.40*
IQ	1.44	0.72	6.04	2.46*	15.47	5.17**

Non-Language Test

Subtest	Control vs. Country Maori (*df* = 256)		Control vs. City Maori (*df* = 173)		Control vs. Market-Garden Maori (*df* = 161)	
	Diff.	*t*	Diff.	*t*	Diff.	*t*
Pict. Comp.	1.19	3.53**	1.20	2.63**	2.72	5.38**
Pict. Arr.	3.07	6.77**	2.79	4.56**	4.63	6.81**
Block Counting	1.95	4.81**	2.01	3.66**	3.26	5.35**
Differences	1.31	3.69**	1.32	2.60**	2.60	3.57**
Series	2.76	5.46**	1.85	2.70**	3.51	4.61**
Total	10.37	6.38**	9.14	4.24**	18.33	7.65**

Note.—In the preliminary analysis of variance, the *F* ratio for Substitutions was not significant; consequently, *t* ratios were not calculated.

* Significant at the .05 level.

** Significant at the .01 level.

homogeneous. With few exceptions, similar findings were obtained with the NLT.

Analysis of ratings was carried out for 697 Maori children of the 11 through 14 age groups for whom scores on both tests and a full set of ratings were obtained. For every subtest and for the total PMA and NLT scores, the pupils of each age group were divided into two groups as nearly as possible equal, according to whether their scores fell above or below the approximate median score of the distribution. The distributions of teachers' ratings among the high and low scoring groups were then compared by means of the chi-square test. The relationships of all five rated variables and each of the total scores, and also those between each of the subtest scores and the ratings of acculturation and educational progress, were assessed in this manner. The findings are given in Tables 4, 5, and 6.

This method was employed because the chi-square values might be summated over the age groups to give an approximate indication of the degree of association of test scores and ratings over the total group of *S*s. Actually, in the case of the subtests in particular, fluctuations in the value of chi square from one age group to another were so considerable that a comparison of total chi-square values would be misleading.

Two of the subtests only, V and W, showed a consistently significant relationship with the acculturation variable over all four age groups; only the Series tests showed a complete absence of any significant relationship. Although in general the subtests of the NLT tended to give fewer significant chi squares, the total score obtained by a simple summation of the scores of the subtests was just as highly related to the acculturation ratings as was the total PMA score.

Two of the PMA subtests, V and R, were significantly related to educational progress at least at the .01 level for each age group; N and Picture Completion were consistently related to this variable at least at the .05 level. Although the PMA subtests appear indi-

TABLE 4

RELATIONSHIPS BETWEEN HIGH AND LOW TOTAL SCORES ON THE PRIMARY MENTAL ABILITIES AND NON-LANGUAGE TESTS AND TEACHERS' RATINGS OF MAORI CHILDREN[a]

(Median test)

Total Primary Mental Abilities Scores *vs.* Ratings by Age Group

Rated variable	11.0–11.11 (N = 153)	12.0–12.11 (N = 160)	13.0–13.11 (N = 191)	14.0–14.11 (N = 193)
School attendance	13.422**	17.031**	5.094	12.953**
Continuity of education	5.204	2.912	7.275*	3.254
Verbal fluency	8.095*	15.108**	23.741**	16.775**
Acculturation	11.461**	16.686**	11.452**	22.425**
School progress	30.451**	16.326**	16.670**	27.547**

Total Non-Language Test Scores vs. Ratings by Age Group

Rated variable	11.0–11.11 (N = 153)	12.0–12.11 (N = 160)	13.0–13.11 (N = 191)	14.0–14.11 (N = 193)
School attendance	15.721**	5.021	17.108**	5.735
Continuity of education	2.945	.583	.808	1.058
Verbal fluency	10.211**	12.796**	9.707*	19.004**
Acculturation	17.202**	10.584**	19.770**	21.921**
School progress	25.900**	16.980**	27.680**	31.138**

* Significant at the .05 level.
** Significant at the .01 level.
[a] Chi-square values (df = 2, for each age group).

TABLE 5

RELATIONSHIPS BETWEEN HIGH AND LOW SUBTEST SCORES AND TEACHERS' RATINGS OF ACCULTURATION OF MAORI CHILDREN[a]

(Median test)

Subtest	11.0–11.11 (N = 153)	12.0–12.11 (N = 160)	13.0–13.11 (N = 191)	14.0–14.11 (N = 193)
Verbal meaning	10.743**	14.760**	13.204**	13.917**
Space	19.330**	4.584	2.020	3.564
Reasoning	6.191*	1.539	9.324**	14.648**
Number	8.442*	0.512	8.348*	4.925
Word Fluency	9.114*	11.961**	6.929*	11.106**
Substitutions	2.745	1.105	4.905	30.360**
Pict. Comp.	4.231	0.638	3.294	11.682**
Pict. Arr.	10.475**	4.896	16.629**	1.860
Block Counting	5.100	0.942	5.965	14.010**
Differences	3.055	3.551	1.326	8.503*
Series	0.699	3.140	5.682	4.844

* Significant at the .05 level.
** Significant at the .01 level.
[a] Chi-square values (df = 2 for each age group).

vidually to be the better predictors of educational progress, the total scores on the PMA and NLT show a very similar and only moderate degree of association with the progress ratings. (Contingency coefficients derived from the total chi squares are .34 and .36 for the PMA and NLT, respectively.)

TABLE 6

RELATIONSHIPS BETWEEN HIGH AND LOW SUBTEST SCORES AND TEACHERS' RATINGS OF EDUCATIONAL PROGRESS OF MAORI CHILDREN[a]

Subtest	11.0–11.11 (N = 153)	12.0–12.11 (N = 160)	13.0–13.11 (N = 191)	14.0–14.11 (N = 193)
Verbal Meaning	28.142**	29.698**	22.957**	16.352**
Space	5.709	8.452*	5.902	5.382
Reasoning	18.313**	9.991**	14.453**	21.164**
Number	20.560**	6.004*	11.258**	18.458**
Word Fluency	21.399**	18.350**	5.120	14.528**
Substitutions	13.772**	9.600**	2.144	26.709**
Pict. Comp.	8.596*	7.428*	8.100*	9.340**
Pict. Arr.	7.893*	3.637	13.877**	5.213
Block Counting	7.246*	8.341*	5.775	12.187**
Differences	10.124**	8.169*	3.299	9.867**
Series	2.085	2.881	12.139**	12.252**

* Significant at the .05 level.
** Significant at the .01 level.
[a] Chi-square value (df = 2 for each age group).

Total scores on both tests were significantly related to school attendance and conversational verbal fluency. Continuity of education, however, appears to be a less potent influence on test scores; the relationship between this variable and total NLT scores was, in fact, negligible.

DISCUSSION

It is evident that the Maori groups differ too greatly among themselves to be considered a single population for comparison purposes in a cross-cultural study of test performance. The significant differences between the means and between the variances of the three Maori groups are evidence to this effect. The findings constitute further evidence that cross-cultural studies of intelligence test performance cannot yield worthwhile conclusions unless socio-economic and educational differences within the different cultural groups are taken into account. They show, too, the futility of drawing general conclusions about the ability of a cultural group from studies of samples taken from a single locality, a practice which has been followed in some prior studies of Maori intelligence. In this connection, it should also be borne in mind that the control group in the present study was drawn largely from an urban area and so was probably slightly above the average for the New Zealand population.

The superiority of the country Maori on the PMA may arise from the fact that pupils in the Maori schools are provided with an educational program which is designed to meet their needs and to allow for handicaps such as poor command of English. The educational advantage of being taught in a Maori school may more than counterbalance the more intense contact with European culture experienced by the city Maori. This is especially likely to be the case when the test in question is for the most part closely related to basic school skills.

On the other hand, it is possible that a process of selective migration may be operating in the opposite direction from that which has been presumed to occur among Europeans and Americans. For economic reasons, the Maori is sometimes forced to seek work in the city where the majority of Maori are unskilled laborers. Perhaps it is the more skillful and alert Maori, who is more ready to learn from experience, who succeeds in farming successfully or holding a steady position of responsibility in his own community.

In addition, Maori children in city areas frequently enjoy a less stable home and community life than those living in country areas; there is thus more likelihood of maladjustment and emotional difficulties which may impede educational progress.

Both of the two last-mentioned factors, selective migration and emotional instability, may contribute to the very poor test performance of the market-garden Maori. It is in the areas in which he lives that the Maori meets with the maximum of discrimination; conditions of housing are poor and standards of living very depressed. The work is largely seasonal and is likely to attract mainly those Maori who cannot find better-paying and more stable employment elsewhere.

No study of selective migration has, to the writer's knowledge, been carried out in New Zealand. The selective migration hypothesis, as it has been applied to explain differences in test performance between northern and southern Negroes, has been severely criticized by Klineberg (1). In view of the fact that the findings about the Maori are the reverse of those generally found for colored subjects in the United States, New Zealand would seem to be a good ground for testing out once more the selective migration hypothesis by experimental investigations. A large number of Maori families return to the country after a period of life in the city; many others move into the city from country areas. What, we might ask, happens to the intelligence test performance of Maori children when the families have made such a move? For the present, however, it is probably safer to look for an explanation of the better performance of the country children in terms of educational, social, and emotional adjustment.

The interpretation of scores for separate factors of the PMA must be guarded in view of the fact that we are dealing with a group that is culturally different from the group on which the factors were isolated. Assuming, however, that the test measures approximately the same set of abilities for Maori and control alike, we may say that while the control group tends to be consistently superior in respect to the Factors V, S, and R, scores for the Factors N and W sometimes favor the Maori, particularly the country group. The finding for N is surprising; while the high scores of the country Maori could reflect to some degree the emphasis placed on basic procedures in Maori schools, the fact that the city group was not significantly inferior to the control group at any age level suggests that this is an ability in which the Maori is potentially strong. The comparatively low Maori score for V and comparatively high score for W seem to bear out a frequently made observation that Maori students tend to be verbally productive and fluent but lack preciseness of expression. The discrepancy undoubtedly reflects the same kind of difficulty that can be seen in the foreigner who has acquired a considerable English vocabulary, but whom the nuances of the language still escape.

The pattern of Maori strengths and weaknesses on the PMA may also afford a clue to the fact that Maori children tend to make relatively poor school progress. They do comparatively poorly on both V and R; yet these appear to be more highly related to school progress than any other subtests, both according to Thurstone and Thurstone (3) and on the evidence of the present study. Therefore, although the mean IQ for the country Maori group is generally above 100 and the mean IQ for the city Maori is generally little below, the subtest patterning suggests that the Maori is likely to perform in school at a level below that which would be expected from his total IQ on the PMA test. The existence of such predictive patterns on other tests and among other groups of *Ss* may perhaps have received too little attention; psychologists are so concerned with detecting potential for better performance than would be indicated by total test scores that they rarely concern themselves with patterns which might lead one to predict poorer school performance than would be expected from the total score or IQ alone.

The results throw a great deal of doubt on the general efficacy of nonverbal tests for the assessment of the intelligence of children who are culturally handicapped. On the whole, results on the NLT proved more unfavorable to the Maori, in comparison with the controls, than did those on the PMA. In addition, on the one PMA test which calls for the handling of perceptual relationships, i.e., S, they also did relatively poorly. The tendency for the more strictly verbal tests to be individually more highly related to acculturation than the remaining tests may be in part explained by the fact that acculturation was defined partly in terms of parental command of the English language. Thus, there may be other cultural differences operating which partly account for the generally poor showing of the Maori on the perceptual tests. In other words, the fact that the verbal tests are within the Maori group more highly related to the acculturation ratings may be due solely to the way in which acculturation was defined by the scale.

If a test is to be used as a predictor of immediate educational progress, obviously even for the Maori children the traditional type of verbal test seems to be the more efficient; in fact, V appears to be about as good a predictor by itself as either of the test batteries as a whole. This is, of course, not surprising, for educational progress is obviously highly dependent upon command of the language of instruction, which in all New Zealand schools is English.

The problem of getting at "intellectual potential" in the case of these Maori children must be considered in the light of their educational background. Compulsory education starts at seven years of age in the Maori schools; consequently, all of our *S*s had had at least four years of schooling with English as the language of instruction. One would expect the more alert and intelligent child to gain more from exposure to educational influences than the less intelligent one. Thus, although on account of bilingualism and inaccuracy in the use of English in their everyday environment they may not have had the opportunity to acquire precision in the use of language, their ability to execute a simple operation such as the addition of numbers with speed and accuracy and the size and availability of their vocabulary, as opposed to accuracy of usage, may provide as good indications of "potential" as their performance on perceptual tasks. Many of the culturally handicapped children who are examined in other countries are in the same situation as the Maori in New Zealand; they have been exposed for some time to educational influences within the dominant culture. Judging from the results of this study, tests such as N and W may be useful for assessing potential under these circumstances. They are probably also more useful predictors of immediate academic success than most perceptual tests; however, as we have pointed out above, they may lead to too high an expectation if tests such as V and R do not produce corresponding results.

It should, finally, be emphasized that the varying performance of Maori children may not be accountable solely in terms of cultural factors. On the PMA, the country group was generally somewhat superior to the city group. Yet, when a comparison of the distributions of ratings of acculturation was made for these two groups, the city group proved to be significantly more acculturated (chi square = 11.889; $p < .01$). Since most of the country group was taken from isolated communities, this is what we should expect. Moreover, the tests on which the Maori children were relatively the most successful, N, W, and Substitutions, appear to be more highly related to the acculturation variable than do others on which they do less well. Even allowing for the fact that our definition of acculturation was probably somewhat inadequate, the findings therefore suggest that we may be dealing with real strengths and weaknesses, which are somewhat independent of cultural factors as here defined, though they may perhaps be partly the outcome of exposure to particular educational methods.

SUMMARY

The SRA form of Thurstone's Test of Primary Mental Abilities and a specially compiled nonverbal test battery were administered to Maori children aged 11 through 15 and to a control group of New Zealand children of European origin. The Maori children

were taken from three areas: (*a*) the city of Auckland and the town of Whangerei; (*b*) semi-rural, market-garden areas, bordering on Auckland; and (*c*) outlying country areas. Wherever possible, teachers' ratings were obtained for Maori pupils for five variables which were expected to show some relationship to test scores.

The Maori groups differ so greatly among themselves that they cannot be considered a single population for purposes of comparison with the control group. The differences between the Maori groups appear to reflect the influence of educational, socioeconomic, and adjustment factors, though a selective migration process of a reverse kind from that alleged to occur among European groups may also be operating.

On the PMA, the Maori appear to be relatively strong on the Factors N and W, and relatively weak on the Factors V, S, and R. This patterning of the subtests may give an indication of why Maori children tend in general to do less well in school than would be expected on the basis of Maori ability as measured by the total test score.

Generally speaking, the Maori groups did less well in comparison with the control on the totally nonverbal test than they did on the PMA. This was consistent with their poor showing on Thurstone's S Test, since the nonverbal battery relied heavily on the discovery or recognition of perceptual relationships. Some doubt is thrown on the efficacy of nonverbal tests for the assessment of ability of culturally handicapped groups.

Total scores on the PMA and nonverbal battery appear to be about equally related to teachers' ratings of school attendance, conversational verbal fluency, and acculturation. The PMA, but not the nonverbal battery, was also related to continuity of education.

The PMA and nonverbal battery were about equally related to school progress when only total scores were considered. However, strictly verbal tests are individually generally the best predictors; V alone appears to predict success about as well as either of the test batteries taken as a whole. Factors N and W show some relationship to educational progress and do not generally seem to put the Maori at a disadvantage but may lead one to expect better progress than the Maori child is, in fact, likely to make. While they may not be good indicators of "potential" for this culturally handicapped group, V and R are probably better indicators of more immediate school progress.

REFERENCES

1. KLINEBERG, O. *Characteristics of the American Negro.* New York: Harper, 1944.
2. REDMOND, M., & DAVIES, F. R. J. *Standardization of two intelligence tests.* Wellington: New Zealand Council for Educ. Res., 1940.
3. THURSTONE, L. L., & THURSTONE, T. G. *Examiner Manual for the SRA Primary Mental Abilities—Intermediate.* Chicago: Science Research Associates, 1948.

Received July 3, 1957.

CRITIQUE AND NOTES

SOME SHORTCOMINGS IN PROJECTIVE TEST VALIDATION

KENNETH PURCELL

University of Kentucky

Psychologists have sometimes taken the position that projective tests have done their work when they reveal the inner, private world of the subject. Occasionally, this point of view makes for a situation in which neither the test data nor the interpretations of these data can ever be explicitly and publicly evaluated. Imagine a situation with which most clinical psychologists doing diagnostic work are familiar: The psychologist makes certain statements about *S*'s characteristics on the basis of projective test data. If, over a period of time, *S* shows no sign of manifesting such characteristics either in daily life or in psychotherapy, the diagnostician has to offer some explanation. He has several alternatives. He may accuse his test data of being unreliable and limited as a sample of behavior. He may re-examine the theories he used in interpretation. Or he may even say that his data and inferences refer to a deeper level of the unconscious than either *S*'s overt behavior or his productions in therapy. This last assertion can, of course, never be proved wrong, and the psychologist may go on merrily describing one unconscious after another, thoroughly insulated from observable fact.

This little bit of caricature highlights the desirability of a validity criterion which includes behavioral prediction. As a matter of fact, however, a number of investigations have utilized such a criterion with inconclusive or negative results being at least as frequent as positive findings. It is the main thesis of this paper that some of the sources of invalidity in projective techniques are avoidable and related to inadequate conceptual schemes.

One of the serious deficiences in recent work is the tendency to ignore a significant aspect of the test data. Studies seeking to relate test performance to overt behavior have commonly failed to consider signs of defensive, inhibitory tendencies. Yet one of the major concerns of clinical psychologists is the study of conflict in human behavior. The kind of conflict that is usually of central importance in personality disturbance is that described by Dollard and Miller (2) as approach-avoidance conflict. Their analysis of approach-avoidance behavior in therapy may also be applied to what occurs in the projective test situation. Suppose an *S* has strong aggressive tendencies (approach) which are inhibited by associated anxiety (avoidance). In testing, as in therapy, an attempt is made to reduce the strength of avoidance, thus permitting the expression of ordinarily forbidden impulses. The avoidance gradient is lowered in two main ways: (*a*) by the permissive, accepting demeanor of the tester, and (*b*) by the deceptive quality of the instructions and the nature of the tests, which tend to prevent *S*'s recognizing the connection between his response and the forbidden impulse.

Henry and Rotter (6) reported a study of situational influences on Rorschach responses which amounted to a manipulation of these approach-avoidance tendencies. They administered the Rorschach to a control group, using the standard Klopfer instructions, but the experimental group was given the added information that the test has been used for detecting and studying serious emotional disturbance. The additional information should theoretically have the effect of keeping avoidance gradients relatively higher in the experimental *S*s. Fewer responses indicating impulse expression or intense conflict should therefore occur in this group. According to Henry and Rotter, the experimental group gave fewer responses, more good form level, more populars and animal responses, more form responses, and fewer aggressive responses.

However, even under standard projective test instructions, avoidance signs are almost never eliminated from the test data. In fact, under certain circumstances, they may be enhanced. A more detailed examination of the Miller conflict model provides a basis for understanding this. Although the model was derived from studies of animals in spatial conflict situations, Miller has suggested that it may usefully be applied to nonspatial conflicts in humans. The concept of an approach and avoidance gradient has been used to describe the overt behavior of an animal with respect to an object simultaneously representing both a goal and a threat. These gradients may be represented schematically as sloping lines on a set of coordinates, the abscissa indicating units of distance from the object (origin of the coordinates) and the ordinate indicating the strength of the approach or avoidance tendency. The tendency to approach or to avoid is maximal in the presence of the object and decreases with distance. The fear gradient is steeper than the goal gradient, which means that as the organism comes closer to the goal, its tendency to run away from it increases more rapidly than does his tendency to approach it. Finally, the strength of the two gradients depends upon drive level. An increase in drive raises the height of the entire gradient. Differences in heights of the gradients determine the point at which the two slopes cross (if at all) and the range on the abscissa characterized by approach and avoidance behavior, respectively.

In the light of this model, how may defensive fear indications actually be increased in performance under projective testing? If avoidance is reduced, but not enough to permit the approach all the way to the goal (the verbal representation of an impulse, motive, or habit tendency), then the impulse will be nearer expression. But the point of intersection of approach-avoidance gradients will be at a higher level than before. Hence, more intense signs of conflict and ambivalence should be evident in *S*'s test responses. Simlarly, the Rorschach test, because it is unusually indirect, probably lowers the avoidance gradient more than, say, the TAT. Therefore, the net approach tendency is likely to be stronger for Rorschach than for TAT responses. Rorschach approach images should be more directly represented. At the same time, more fear and conflict ought to be elicited because *S* is nearer the goal, and the strength of both tendencies is increased. As Miller points out, these deductions will hold true only for the range within which the approach-avoidance gradients intersect.

On theoretical grounds, then, one expects an abundance of projective test data expressive of fears and inhibitions. Those clinicians who regularly deal with projective test protocols can readily testify to this not very remarkable deduction. Nevertheless, a significant portion of the validity studies concerned with behavioral correlates of projective performance seem to ignore avoidance behaviors as criteria. Smith and Coleman (13), for example, predicted a curvilinear relationship between hostile content in the Rorschach and overt behavior. Their hypothesis was based on the hydraulic conception of an inverse relationship between overt behavioral discharge and fantasy. A low but statistically significant correlation was found between ratings of overt hostile behavior and Rorschach content. *S*s high in Rorschach hostility scores expressed little overt aggression in a classroom. The authors suggested that "*S*s evidencing low hostile content scores could theoretically have been *S*s who discharged tension immediately and regularly or who tended to be impunitive and to show little evidence of hostility arousal in their everyday behavior." They chose the latter possibility because their *S*s low in Rorschach hostile content actually revealed little overt hostility in the classroom. However, evaluation of the relative strengths of approach-avoidance tendencies along the relevant hostility dimension would permit more precise prediction. For one thing, it would not be necessary to entertain alternate hypotheses for the group low in Rorschach hostile content and then to select the correct hypothesis on a *post hoc* basis. A single prediction could be made for every individual, based on approach-avoidance ratios.

Finney (4) tested the hypothesis that response content in the Rorschach can be related to assaultive behavior. He found the relationship between a Destructive Content scale and assaultive behavior ratings to fall just short of the 5 per cent level of significance in spite of $N = 117$. Here, again, there was an absence of any attempt to conceptualize the avoidance gradient variables and to make a behavioral prediction in terms of an impulse-control balance. As a matter of fact, there seems to have been a lumping together of approach and avoidance content within the same scoring scale. The Possibly Destructive category, for instance, included responses in which the "concept is more likely than not to attack, injure, or destroy something" as well as responses in which the concept was considered "frightening or dangerous."

Similarly, Stone (14), in constructing a TAT Aggressive Content scale, indicated that one of his primary aims was to relate it to overt behavior. Yet nowhere in his scoring system does he appear to make any provision for the defensive avoidance forces which are of such importance in modifying or inhibiting the behavioral resultant of impulses.

In concluding that amount of hostility displayed by a patient in the TAT is not by itself a reliable index for predicting overt hostile behavior under stress, Gluck (5) noted the desirability of considering other factors, such as anxiety associated with hostility and direction of hostility. Studies which have attempted to take these variables into account (11, 12) seem to have been more successful in relating fantasy material to overt behavior than those mentioned above.

Research by McClelland and his associates (10) has led to a distinction between two aspects of achievement motivation, hope of success and fear of failure. As Clark and others (1) have pointed out, this amounts to a distinction between an approach motive and an avoidance motive. Clark found that, with regard to a level of aspiration measure, students at the extremes of the aspiration continuum have lower n-Achievement scores than students in the middle of the continuum. However, when n-Achievement scores were separated into positive goal imagery and deprivation imagery categories, it became clear that the n-Achievement score of the middle group was overwhelmingly a function of deprivation imagery.

Thus far, the discussion of the Miller conflict model has been confined to those circumstances in which there is assumed intersection of approach-avoidance gradients prior to reaching the goal. Where the gradient of avoidance is so weak that it no longer intersects the gradient of approach, a situation exists which may have relevance to another problem. This has to do with the rather widely held hypothesis that fantasy has a substitutional or compensatory relationship to overt behavior. Although commonly held, this view has been questioned by some investigators. McClelland *et al.* have been among the doubters on the basis of inferences from TAT fantasy data. Purcell (12) suggested that certain individuals may actually be stimulated to overt action by engaging in fantasy. This suggestion was made because of the finding, contrary to the fantasy substitution hypothesis, that antisocial individuals produced more aggressive fantasy than nonantisocial neurotics. Antisocial persons generally have very weak avoidance tendencies with respect to aggression. It may be that under circumstances in which the approach tendency is so strong relative to avoidance that no intersection occurs, fantasy tends to become a preparatory stimulus to action rather than a drive-reducing substitute. The reason for this hypothesized differential role of fantasy may be that persons high in approach and low in avoidance have not been faced with the need for learning to utilize fantasy in a substitutive capacity because behavioral discharge was readily accessible.

Feshbach (3) found that an insulted group of students who had an opportunity to express aggression in fantasy displayed significantly less aggression toward the experimenter who provoked hostility than did the control groups who engaged in nonfantasy activities. Presumably these students were, in the main, individuals with fairly well developed inhibitions (high avoidance gradients) in the aggressive sphere. Perhaps the same experiment repeated with antisocial individuals would yield quite different results. Or, were the *S*s divided into two groups, one having relatively high and the other relatively low approach-avoidance ratios, fantasy might be found to function more effectively as a drive-reducing agent for the latter.

A basic question, of course, is that of how to measure the strength of approach-avoidance tendencies in projective test responses. Numerous scoring schemes have been proposed for the TAT and Rorschach. In selecting measurement indices for the two tendencies, it may be well to keep in mind such established indices of habit strength as probability of response evocation, latency, and amplitude or intensity (7). These suggest corresponding measures in the test data. Two common measures might be used to reflect probability of response evocation: frequency of a thematic or imagery type in a given protocol and the relative novelty or idiosyncratic quality of an image or theme in terms of normative data. Latency might be evaluated by the initial appearance of a type of theme with reference to time and to position in the response sequence. Scaling of amplitude or intensity would at first require a more qualitative judgment regarding the content form in which a given habit tendency or impulse is couched. It is evident that similar content may be produced by two persons in ways that suggest important variations in the kind of control exercised over the impulse represented.

Another important conceptual omission in this field has to do with the notion of generalization. Behavioral prediction is likely to be accurate to the extent that the variables present in the test situation are similar to those in the prediction situation. This may include not only the characteristics peculiar to the tester—subject relationship—but also the actual content of the test stimuli, e.g., the kind of figures shown in the TAT pictures. For example, an experimental test validation that required the prediction of reaction to frustration among adults would probably be doomed to mediocre results at best without a more precise definition of the problem. A person's behavioral reaction to frustration is probably in large part a function of the frustrating agent and the situational context. For some individuals, depending upon whether the frustration was induced by a child, a peer, or a superior, the reaction may be one of overt irritation and anger, sullen withdrawal, or an outwardly compliant, submissive response. It is not enough simply to seek to relate projective data to, say, hostile behavior. The ideal to be pursued is an exact statement of the nature of the hostile behavior criterion—hostile to what degree, how expressed, toward whom, and under what circumstances. Then, for specific behavioral predictions, stimulus materials can be meaningfully selected or constructed so as to have them more closely resemble the predictive situation which constitutes the criterion.

Kagan (8) illustrated this point of view in an experiment demonstrating that frequency of fighting themes in TAT-like pictures is more directly related to overt fighting behavior than to other types of fantasy aggression. He constructed 13 pictures especially and specifically designed to elicit relevant responses. On the other hand, Gluck (5), in his study of provocation of hostile behavior under stress, seemed to overlook completely the implications of a generalization gradient between the test and predictive situations. The stress in this experiment was a "hostile, domineering authority figure." Yet the TAT cards used were 1, 11, 10, 8GF, 13MF, and 18GF. Pictures which might have elicited the most meaningful information on the basis of situational similarity, e.g., 7BM, 12M, and 8BM (cards in which male authority figures are commonly projected), were not used. Perhaps also pertinent here is one result of the Smith and Coleman (13) investigation already cited. They reported that the only Rorschach content relationship significant at the .01 level was that between Rorschach hostile content and physical hostility. It would be of interest to know how much of the content involved percepts of physical hostilities and whether an analysis similar to Kagan's would yield a similar result.

The results of this brief review and analysis of some of the literature strongly support Korner's (9) advice to distinguish the real from the avoidable sources of invalidity in the field of projective techniques by a more careful theoretical analysis.

SUMMARY

This paper has sought to demonstrate that the negative or inconclusive results obtained in many validity studies of projective techniques are, in part at least, determined by important defects in the experimental conceptualization of the problem. The major deficiencies noted were (*a*) a tendency to ignore the significant avoidant aspects of behavior as reflected in the test data, and (*b*) a failure to recognize the importance of the generalization gradient from test to predictive situation as a relevant variable affecting predictive accuracy. An hypothesis was offered regarding a differential role of fantasy in relation to overt behavior. Finally, certain concepts from general behavior theory, particularly the Miller conflict model, were used in formulating the ideas expressed.

REFERENCES

1. Clark, R. A., Teevan, R., & Ricciuti, H. N. Hope of success and fear of failure as aspects of need for achievement. *J. abnorm. soc. Psychol.*, 1956, **53**, 182–186.
2. Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
3. Feshbach, S. The drive-reducing function of fantasy behavior. *J. abnorm. soc. Psychol.*, 1955, **50**, 3–11.
4. Finney, B. C. Rorschach test correlates of assaultive behavior. *J. proj. Tech.*, 1955, **19**, 6–16.
5. Gluck, M. R. Relationship between hostility in the TAT and behavioral hostility. *J. proj. Tech.*, 1955, **19**, 21–26.
6. Henry, E. M., & Rotter, J. B. Situational influences on Rorschach responses. *J. consult. Psychol.*, 1956, **20**, 457–462.
7. Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.
8. Kagan, J. The measurement of overt aggression from fantasy. *J. abnorm. soc. Psychol.*, 1956, **52**, 390–393.
9. Korner, Anneliese F. Limitations of projective techniques: Apparent and real. *J. proj. Tech.*, 1956, **20**, 42–47.
10. McCelland, D. D., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
11. Mussen, P. H., & Naylor, H. K. Relationships between overt and fantasy aggression. *J. abnorm. soc. Psychol.*, 1954, **49**, 235–240.
12. Purcell, K. The TAT and antisocial behavior. *J. consult. Psychol.*, 1956, **20**, 449–456.
13. Smith, J. R., & Coleman, J. C. The relationship between manifestations of hostility in projective tests and overt behavior. *J. proj. Tech.*, 1956, **20**, 326–334.
14. Stone, H. The TAT aggressive content scale. *J. proj. Tech.*, 1956, **20**, 445–452.

Received May 20, 1957.

SOME EFFECTS OF INVOLVEMENT UPON EVALUATION[1]

HAROLD B. GERARD

Bell Telephone Laboratories, Murray Hill, N. J.

The general hypothesis being investigated here concerns the relationship between involvement in a group activity as it affects the evaluation of the group members' contribution; the greater the involvement, the greater are the predicted effects upon evaluation. Involvement was varied by superimposing two experimental manipulations. Access to the major activity of the group was varied and so was the degree to which the activity had potential motivational consequences.

METHOD

Subjects. The 36 individuals who comprised nine four-person experimental groups were students in investigator's introductory psychology class at New York University. Each of the experimental groups was composed either entirely of men or entirely of women.

The experimental setting. A few days before the experiment, the class members were asked to be prepared to discuss the chapter in their text on intelligence tests. On the day of the experiment the class was divided into groups of four persons each. They were told to discuss the material within the framework of several general questions concerning intelligence tests with the aim of formulating some specific questions which they would like to put before the entire class at the next class meeting. Individuals were assigned to the groups so as to minimize the number of acquaintances.

---

[1] The writer wishes to thank Daniel Wilner for his help in running the experiment. Thanks are also due Ruby Weinberg for help with the analysis and to Morton Deutsch for suggestions and criticism. This study which was conducted at the Research Center for Human Relations, New York University, was sponsored by the Office of Naval Research, Contract NONR 285 (10). The United States Government is authorized to reprint this article in whole or in part.

By a random process, one individual was assigned to the role of recorder. The recorder was told that he could not participate in the discussion but should take notes on what was said. The other three members were the active individuals whom we shall henceforth call the discussants. A group discussion ensued which lasted for thirty minutes.

The experimental conditions. Accessibility to activities should produce high involvement only if the activities are seen as motivationally relevant, i.e., if engaging in the activity has consequences for the individual. The experiment was designed to produce two degrees of motivational relevance. *S*s in four of the groups were told that the discussions would be carefully observed by the experimenter and the quality of their performance would in part determine the grade they would receive in the course. *S*s in the other five groups were not given these instructions. In this latter condition of low motivational relevance, involvement is allowed to vary as a function of numerous chance factors operating in the situation, whereas in the high condition a systematic heightening of involvement is expected. The treatments were run in adjoining rooms.

The questionnaire. After the discussion, each *S* ranked the group members concerning certain performance criteria:

1. Who knows the most.
2. Who explained things the best.
3. Who contributed the most.

They were also asked to predict the rankings made by each of the other participants on all three criteria.

RESULTS

The rank order correlation coefficients between the orderings *S* predicted each of the others would make and the orderings each actually made were computed. Since the recorder did not participate in the discussion, his rank placement in all of the orderings was eliminated. His rankings of the others, however, were included. The average rho for the predictions each *S* made of the others' rank orderings on the three performance criteria was then computed. Thus, there were nine rhos for each *S*, one for each of the rank orderings made by the three others in his group on each of the three aspects of group process which were then averaged[2] to obtain a measure of his accuracy of predicting the others' evaluations.

TABLE 1

MEAN ACCURACY OF PREDICTION BY THE RECORDER AND DISCUSSANTS UNDER HIGH AND LOW MOTIVATIONAL RELEVANCE

Condition	Recorder	Discussant
High motivational relevance	.79	.40
Low motivational relevance	.72	.71

TABLE 2

AVERAGE RANK[a] FOR DISCUSSANTS ON THE THREE PERFORMANCE CRITERIA

Criterion	High Motivational Relevance			Low Motivational Relevance		
	Self Rank	Ranked by Others	Expected Rank	Self Rank	Ranked by Others	Expected Rank
Knowledge	2.00	2.46	2.19	2.29	2.40	2.53
Contribution	1.75	2.08	1.83	1.80	2.19	2.07
Explanation	1.92	2.34	1.98	2.20	2.31	2.29

[a] The lower the number the higher the rank.

In Table 1, these indices are presented separately for the recorder and the discussants under high and low motivational relevance. The recorder-discussant difference under high motivational relevance is quite striking ($p < .003$ by the sign test),[3] whereas no such difference appears under low personal relevance. A trend is also indicated between discussants under high and low motivational relevance ($p < .15$ by t).

How can we best account for the relative inability of the discussants under high motivational relevance to predict the rank orderings correctly? The figures in Table 2 indicate the degree to which discussants overestimated their productive value to the group.[4] Since the data for the recorder are based on too few cases to be statistically reliable, they are not presented. The self-rating figures indicate the average rank placement an individual gave to himself. The ratings of others indicate the average rank placement given him by the other group members. The figures show a statistically significant tendency ($p < .05$ by t) for discussants under high motivational relevance to over-estimate their value to the group, whereas no statistically significant difference exists under low relevance. These t tests were performed by comparing, for each *S*, his mean self placement on the three criteria with the mean placement of him by the others.

---

[2] These averages are based upon z-score conversions.

[3] It seemed advisable to test the difference between recorder and discussant accuracy by means of the sign test. The test involved comparing, in each experimental group, the accuracy of prediction of the recorder with each of the three discussants.

[4] Since the recorder was not included in computing the average rhos presented in Table 1, data for the recorder are omitted in Table 2. It is somewhat meaningless to expect the recorder to evaluate his own performance or to have any expectations at all concerning how others had evaluated him since he did not participate. During the questionnaire administration, *S*s encountered considerable difficulty in this respect. The questionnaire should have been constructed so as to eliminate evaluations of the recorder's performance.

In order to determine whether or not motivational relevance in this situation affected how an individual thought the others in the group valued him, the average rank the individual expected others to assign him was computed and compared with the rank they actually gave him. The data on expected rank are presented in the third column of Table 2. Once again overestimation occurs only under high motivational relevance ($p < .05$ by t).

The discussants' failure under conditions of high motivational relevance to predict the relative evaluations placed upon the membership by others (Table 1) can therefore perhaps be attributed to the fact that, as a result of their desire to do well, they tended to overestimate their value to the group. When an individual has a personal stake, as it were, in the quality of his own performance, his judgments of his performance tend to be equilibrated with his expectations. It is as though the judgmental process itself serves, in part at least, to realize one's desires.

SUMMARY

Four-person discussion groups were created in a classroom in which three members were active participants and one member functioned in the peripheral role of recorder. In four of the groups, the discussion occurred under high motivational relevance (a personal stake was introduced) whereas in five of the groups the discussion occurred under low motivational relevance (lack of an explicit personal stake).

After the discussion, each *S* was asked to rank order the membership of his group concerning a number of aspects of group process and to guess the rankings made by each of the others in his group. The data revealed that the recorder-discussant difference in accuracy of prediction occurred only under high personal relevance. Further analysis was undertaken which demonstrated that this differential accuracy can be accounted for by the tendency for discussants under high motivational relevance to over-value their own performance.

Received May 14, 1957.

TASK DIFFICULTY AND CONFORMITY PRESSURES[1]

JANET FAGAN COLEMAN, ROBERT R. BLAKE, AND JANE SRYGLEY MOUTON

University of Texas

ARE conformity pressures more easily exerted when *S*s are less adequately informed regarding the correct response? A direct relationship seems self-evident. Yet, when judgments were made easy or difficult by changing rate of clicking for a metronome counting task, conformity to social pressures was found to be unrelated to "difficulty" (5). Where the objectively correct judgments were rendered "easy" or "hard" by varying the relative lengths of the standard and comparison lines, the relationship was not confirmed in a study with adults (1) though it was with children (2). On the other hand, susceptibility has been reported as a function of difficulty for arithmetic problems (5, 6). While susceptibility may vary directly with difficulty, results from experiments reported do not consistently demonstrate the validity of the relationship. However, none of the tasks described above are subjected to conformity pressures under ordinary social conditions.

The following experiment evaluates the relationship with respect to kinds of information encountered in social situations of everyday life. The materials judged included questions about current events, geography, government, literature, language, and science. Since an individual is expected to be informed about such questions of information, the existence of a relationship between susceptibility and difficulty should be most easily demonstrated for these kinds of materials.

METHOD

The experiment was presented to each of the sixty *S*s as a comparative study of college students today with those twenty years ago in regard to ability to deal with questions involving general information.

Description of the Test Situation

Each participant understood that to save time he was one of four who would be working together. Simulated group procedure was used to create a standard influence situation (3, 4). Fifteen of the men and fifteen of the women responded to each item after hearing the

---

[1] The study reported here was made possible by a research grant from The Hogg Foundation for Mental Hygiene, The University of Texas, Austin, Texas.

TABLE 1
ITEM CONTENT, DIFFICULTY LEVEL, AND PER CENT OF *S*s CONFORMING TO THE INCORRECT ANSWER

Item	*Per Cent Difficulty*[a]		*Per Cent Conforming*	
	Men	Women	Men	Women
1 Knight is the name of the Governor of: Iowa, Indiana, California*[b], *Maine*[c]	10	50	33	60
2 The smallest state in the union is: Connecticut, *Delaware*, New Jersey, Rhode Island*	08	06	27	20
3 Staunch means the same as or the opposite of: *Easy*, Unwavering*, Uneven, Stupid	28	13	27	37
4 The chief justice of the Supreme Court is: Holmes, *Burton*, Warren*, Clark	00	30	20	27
5 Stalwart means the same as or the opposite of: *Tall*, Weak*, Wild, Agitated	22	36	43	40
6 Topsy was: Arabian, Eskimo, *Indian*, Negro*	54	14	40	33
7 The opposite of militarism is: *Nationalism*, Nihilism, Pacifism*, Communism	00	30	37	56
8 Morose means the same as or the opposite of: Urgent, *Docile*, Gloomy*, Thorough	08	10	27	13
9 Avid means the same as or the opposite of: Eager*, *Vivid*, Arid, Volatile	54	33	50	40
10 A ward is: A Public Park, a Political Division of a City*, *A Seat of County Government*, The Business District of a City	00	40	40	50
11 A gene is a structure believed to be responsible for: Resistance to Infection, Inheritance of A Trait*, *The Formation of Gametes*, The Development of Scurvy	02	00	13	10
12 Irksome means the same as or the opposite of: Mortal, Possible, Tiresome*, *Inferior*	30	10	46	13

[a] Ten per cent difficulty means that 10% of the standardizing group gave an incorrect response.
[b] The correct answer for each item is indicated by an asterisk (*).
[c] The response given by the three others prior to *S*'s response is italicized in each case.

reports of three men. The other fifteen men and fifteen women responded after three women.

Task

The task involved twelve items embedded among a total of 35 questions with instructions to choose the correct answer from four alternatives given. The first three individuals who responded gave an incorrect answer to the 12 items as shown in Table 1. They gave both correct and incorrect responses for the 23 noncritical items. An answer was scored as a conforming response if it was the same as that by the three persons responding before the *S*.

Selection of Items

Level of difficulty was based on the reactions to the items by 50 men and 50 women answering under private conditions. Items were selected, for another part of the study, so that some were better known by men than women and vice versa. Judged in terms of percentage of the standardizing group giving the correct answer, difficulty ranged from zero to 54 per cent. Since the standardizing group and experimental groups are comparable in terms of age, academic level and major, more frequent yielding under experimental conditions on difficult than on easy items would establish that conformity pressures are more easily exerted when the *S* is less well informed regarding the correct answer.

RESULTS AND DISCUSSION

The frequency of conformity is shown in Table 1, for men and women. The relationship between the difficulty of the items and frequency of conformity is evaluated by rank-difference correlations. The correlation for men is .58, significant beyond the .05 level, and for women, .89, significant beyond the .01 level. Although the sex of the others was varied systematically, there was no interaction between sex and level of difficulty.

For both men and women, therefore, conformity pressures are more easily exerted when *S*s are least well informed and the task involves items of general information. The results are interpreted as indicating that if an individual is certain of the correct answer, he is more able to resist pressures being exerted by being more able to respond in terms of internal cues. Supplemental external information as the basis of his response is more frequently employed when the person is unfamiliar with the correct answer. A direct implication of these findings is that susceptibility to conformity pressures can be decreased by increasing an individual's ability to make a competent selection of responses.

REFERENCES

1. Asch, S. E. Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men.* Pittsburgh: Carnegie Press, 1951.
2. Berenda, Ruth W. *The influence of the group on the judgements of children.* New York: King's Crown Press, 1950.
3. Blake, R. R., & Brehm, J. W. The use of tape recording to simulate a group atmosphere. *J. abnorm. soc. Psychol.*, 1954, **49**, 311–313.
4. Blake, R. R., & Helson, H. (Eds.). *Adaptability screening of flying personnel. Situational and personal factors in conforming behavior.* School of Aviation Medicine, USAF, Report No. 56-86, September, 1956.
5. Blake, R. R., Helson, H., & Mouton, Jane S. The generality of conformity behavior as a function of factual anchorage, difficulty of task, and amount of social pressure. *J. Pers.*, 1957, **25**, 294–305.
6. Coffin, T. E. Some conditions of suggestion and suggestibility: A study of certain attitudinal and situational factors influencing the process of suggestion. *Psychol. Monogr.*, 1941, **53**, No. 4.

Received June 7, 1957.

"FAULTY" COMMUNICATION AND THE SPIRAL AFTEREFFECT: A METHODOLOGICAL CRITIQUE[1]

EUGENE S. GOLLIN

Queens College

AND NORMAN BRADFORD

University of Minnesota

Inability to report aftereffect following fixation of a rotating Archimedes spiral has recently received considerable attention by investigators interested in utilizing this failure as a diagnostic indicator of brain damage (e.g., 1, 3). The same technique has been utilized in a developmental study by Harding, Glassman, and Helz (2). In this study, no child under CA 55 months or under MA 60 months achieved the success criterion of three out of four trials. These data are interpreted as a possible reflection of neurological immaturity, and a functional similarity between children and adults with organic cerebral damage is alleged. Recognizing that their results may not represent a perceptual failure but rather a failure of report, Harding and his associates *remain convinced* of the validity of the verbal responses. Since the presence or absence of perception of the spiral aftereffect is implicitly or explicitly related to structural and functional characteristics of the cerebral cortex by almost all investigators who have employed the spiral technique, the validity of verbal report is a matter of some consequence. Some *S*s who fail to meet the criterion of success, especially young children and brain-damaged adults, may, in view of the nature of the aftereffect, be unable to supply the verbal designators necessary for the achievement of a positive score.

The aftereffect, depending upon the direction of rotation of the spiral, consists of either a phenomenally receding and/or contracting visual field, or an approaching and/or expanding field. The usual method of obtaining the aftereffect report is to have *S* fixate a rotating spiral (rotation speed varies among investigators, but 78 RPM is fairly standard) for approximately thirty seconds and then report his visual impression when the rotation is abruptly terminated. If during the rotation period the spiral appears to be expanding or approaching, upon cessation of rotation the phenomenal field appears to be receding or contracting and vice versa.

The aftereffect is not confined to the spiral per se. Following fixation of the rotating spiral for a sufficient time, virtually any object in the surround may be employed for the elicitation of the verbal report. This fact widens the methods by which the phenomenon may be investigated.

No attempt has been made, so far as is known, to determine the verbal designators which *S*s employ when confronted with stimulus situations which involve *actual* rather than illusory expansion and approach, and actual contraction and recession. If such designators were available before *S*s were tested in the aftereffect situation, the value of verbal report as a reflection of phenomenal experience would be considerably enhanced.

Method

An Archimedes spiral of 3½ turns (1260°) with an 11½-inch diameter served as the stimulus disc. It was mounted on a phonograph turntable which could be driven at various speeds and reversed by a switching arrangement. The turntable was presented to *S*s at a distance of 10 feet, perpendicular to the line of sight. The rotation speed was 78 RPM and fixation time was 30 seconds.

---

[1] This research was completed while the senior author was at the University of Minnesota. It was supported by a grant from the Graduate School of the University.

Large white balloons with spirals of black ink painted on the surface opposite the nozzles were prepared. These balloons were used in the pre- and postfixation periods as follows: *S* viewed an inflating, deflating, and fixed balloon. Inflation was accomplished by means of compressed air. The hissing noise accompanying the inflation was also maintained during the deflating and fixed presentations. *S* was given a period of exploration of the equipment, and rapport was established. *E* then presented the stationary balloon spiral and said, "See this design? It's called a spiral. See how it goes round and round? Can you say spiral? I have some balloons with a spiral on each one. I'm going to show you one balloon at a time. I want you to tell me each time if the spiral is *getting bigger*, or if the spiral is *getting smaller*, or if the spiral is staying the *same size*." *S* was then seated directly in front of the mounted disc spiral which was covered to prevent distraction during the initial procedure. The directions, "Remember, each time, tell me if the spiral is *getting bigger*, or if it is *getting smaller*, or if it is *staying the same size*," were repeated. The compressed air was released into a deflated balloon held in the position of the mounted spiral and the instructions, "Tell me about the spiral now," were repeated several times if necessary. The air was then slowly released from the balloon and the instructions repeated, "Tell me about the spiral now." Finally, the stationary balloon was presented with the same instructions.

The mounted spiral was then set in motion and uncovered with the following instructions: "Now I want you to look at this spiral. See this point right here (*E* points)? Look right at this point until I tell you to stop. Don't take your eyes off the point." (Continual encouragement was given to keep looking at the center point.) At the end of the fixation period, the stationary balloon with spiral was quickly placed in front of the mounted spiral (obscuring it) with the instructions "Tell me about *this* spiral now."

Each *S* was given four trials in the order A B B A or B A A B. A toy was given between trials to allow a time lapse in order to avoid perseveration and to maintain interest. *S* passed if his report was correct on three of the four trials.

Thus, the balloons with spirals painted upon them served in the prefixation period to elicit verbal designators which accompany an actual approaching and expanding or contracting and receding spiral, and in the postfixation period the stable balloon was used to elicit the illusory aftereffect response. The rotating spiral itself was employed only for fixation purposes.

The *S*s were 23 children with a chronological age range of 38 to 63 months, mean CA = 49.65 months, *SD* = 7.55 months. Mental ages ranged from 42 to 88 months, mean MA = 63.13 months, *SD* = 13.77 months. *S*s were drawn from the nursery school of the Institute of Child Welfare, University of Minnesota.

RESULTS AND DISCUSSION

Seventeen of the twenty-three *S*s achieved the success criterion (at least three out of four correct responses) by utilizing the correct verbal designators when they viewed the spiral painted on the unmoving balloon, after they had fixated the rotating disc. In the Harding, Glassman, and Helz study there were 40 *S*s in the CA range 46 through 63 months of whom 26 failed to achieve the success criterion. In the present study, only one of 17 *S*s in the same age range failed.

Similarly, in the Harding, Glassman, and Helz study, there were 29 *S*s in the MA range 52 through 84 months. Twelve of these *S*s failed. In the present study, only two of 15 *S*s in this MA range failed.

The youngest *S* in the earlier study to achieve success was CA 55 months. In the present investigation, the youngest successful *S* was CA 45 months. The lowest MA at which passing occurred in the earlier study was 61 months. In the present study the lowest passing MA was 48 months.

It is particularly noteworthy that five of the six *S*s who failed in the present study were also unable to respond correctly in the pretest situation, where they were required to respond to actual rather than illusory changes. It may be that failure to report illusory changes among some *S*s is based on the unavailability of the necessary verbal designators rather than upon perceptual deficit.

The balloon technique permits determination of verbal report availability prior to fixation of the rotating spiral, and the posttest elicitation of response by means of a balloon which remains stable with regard to size and motion appears to evoke correct responses from children younger in both CA and MA than previous results suggest.

SUMMARY

Twenty-three children were tested for spiral aftereffect under a method designed to obtain their responses under actual as well as illusory conditions. It was found that virtually all *S*s who responded correctly under actual conditions were able to report correctly under illusory conditions. *S*s considerably younger in CA and lower in MA were able to achieve success in the task under present conditions than were able to achieve success in a previously reported investigation.

REFERENCES

1. GALLESE, A. J., JR. Spiral aftereffect as a test of organic brain damage. *J. Clin. Psychol.*, 1956, **12**, 254–258.
2. HARDING, G. F., GLASSMAN, S. M. & HELZ, W. C. Maturation and the spiral aftereffect. *J. abnorm. soc. Psychol.*, 1957, **54**, 276–277.
3. PRICE, A. C., & DEABLER, H. L. Diagnosis of organicity by means of spiral aftereffect. *J. consult. Psychol.*, 1955, **19**, 299–302.

Received July 1, 1957.

SOCIAL DESIRABILITY RATINGS OF PERSONALITY VARIABLES BY NORWEGIAN AND AMERICAN COLLEGE STUDENTS

O. IVAR LÖVAAS[1]

University of Washington

THE purpose of this study was to find out the degree of relationship between social desirability judgments made by Norwegian and American college students on a set of statements about behaviors that a person might display. It has been found in previous research (4) that high school students from different socioeconomic classes show high agreement in their social desirability judgments of the personality statements investigated in this study. A similar high relationship was found between the judgments of American and Japanese-American college students (3). These correlations are all in the nineties. The scaling procedure in these studies was the same as in the study here reported.

The statements in Edwards' Personal Preference Schedule (PPS) were used. This schedule consists of 135 statements about behavior grouped into 15 personality variables, 9 statements defining each variable. These variables are described as manifest needs by Murray (5) and are labelled Achievement, Deference, Order, Exhibition, Autonomy, Affiliation, Intraception, Succorance, Dominance, Abasement, Nurturance, Change, Endurance, Sex, and Aggression. For a more complete description of this inventory, see Edwards (2).

The Norwegian translation of these statements was checked by faculty members of the Scandinavian Department of the University of Washington and of the Psychology Department of the University of Oslo. The translated list of statements was presented to senior classes at two gymnasia in Oslo in the fall of 1954. The Norwegian sample consisted of 86 *S*s, with a mean age of 17 years. The American sample (see Edwards [1]) contained 152 college students, with a mean age somewhat higher than the Norwegian sample.[2] The statements were presented one at a time and the *S*s rated the statements on a nine-point scale ranging from extremely socially desirable to extremely socially undesirable. The method of successive intervals was used to obtain the scale values of the items.

The product-moment correlation between the scale values of the Norwegian and the American sample was .78. This correlation, which indicates a high agreement with 61 per cent of the variance being common, is probably an underestimate of the true relationship, since errors in translation would attenuate the obtained value.

Thus, it is possible that the remaining 39 per cent of the variance is made up to an unknown degree of error variance due to changes in the statements through translation. However, the assumption was made that chances are 50–50 of one group rating a statement higher or lower than the other group in terms of social desirability. Working on this assumption, consistent differences were found between the two groups in their rating of statements pertaining to 4 of the 15 variables. (This difference, significant at the .05 level, was evaluated in terms of the sign test for paired observations and the binomial expansion, where $p = .5$, $q = .5$, and $n = 9$.) Americans rated the statements pertaining to *Order*, *Intraception* and *Abasement* as more socially desirable than did the Norwegians. The Norwegians, on the other hand, rated the statements pertaining to *Aggression* as more socially desirable than did the Americans.

The personality variables employed in this study have the advantage of being clearly defined as specific behavior is described by each statement. However, specificity limits one's information to only a few facets of a trait. A trait like aggression has many ways of expressing itself in behavior, and only nine rather overt expressions of this trait were employed in this study. Thus the difference on Aggression might disappear or reverse itself if the two groups had rated more subtle expressions of this trait.

The author feels that the paper-and-pencil test approach, even with a translated version of the test, is feasible for cross-cultural studies—provided the test points to specific observable behaviors, as does the PPS. An approach like this throws some light on the similarities and differences of behaviors that are acceptable in the two cultures.

REFERENCES

1. EDWARDS, A. L. The relation between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93.
2. EDWARDS, A. L. *Edwards Personal Preference Schedule. Manual.* New York: Psychological Corp., 1954.
3. FUJITA, BEN. An investigation of the applicability

---

[1] The author expresses appreciation to Allen L. Edwards of the University of Washington and Per Saugstad of the University of Oslo for their help in this study.

[2] Klett (4) found that there were no significant differences in social desirability ratings at these age levels.

of the Edwards Personal Preference Schedule to a cultural sub-group, the Nisei. Unpublished master's dissertation, Univer. of Washington, 1956.

4. KLETT, J. C. A study of the Edwards Personal Preference Schedule in relation to socio-economic status. Unpublished doctoral dissertation, Univer. of Washington, 1956.

5. MURRAY, H. A. *Explorations in personality*. New York: Oxford, 1938.

Received July 17, 1957.

RELATIONSHIP OF MANIFEST ANXIETY TO STIMULUS GENERALIZATION[1,2]

ROBERT E. FAGER AND IRWIN J. KNOPF

State University of Iowa[3]

MANY INVESTIGATIONS have been concerned with relationships between the Taylor Manifest Anxiety Scale (MAS) and a variety of learning phenomena. In the area of stimulus generalization (SG), Rosenbaum (4) found greater responsiveness to generalized stimuli in a spatial situation for high MAS college students than for a low MAS group, but only when the *S*s were given strong intermittent shock during their performance. Using a temporal generalization paradigm, Wenar (6), however, reported greater responsiveness for high MAS group for weak shock and buzzer conditions. Buss (2) trained psychiatric patients to make a verbal response to wooden blocks of a given height. Controlling for directional effects and measuring generalization to blocks of graded heights, he found no difference between high and low MAS groups in their generalization gradients.

In view of these somewhat inconsistent findings and the suggestion that situational factors may account for differences in MAS results (5), the present study was designed to examine the relationship between different levels of anxiety and stimulus generalization in psychiatric patients in a situation different from Buss's.

METHOD

Subjects. One hundred sixty-nine patients, representing a wide variety of psychiatric conditions, were given the Taylor Manifest Anxiety Scale. All were at least 16 years old, were not on any drugs or physical therapy, and scored less than 8 on the *L* scale of the MMPI. They were about evenly distributed by sex and had a median age of 31.2 and an educational level of 12.4 years. Those in the highest (scores of 35 and above) and lowest (17 and below) fifths of the MAS distribution were selected as the high anxiety (HA) group ($N = 34$) and the low anxiety (LA) group ($N = 31$).

Apparatus. The apparatus was similar to that developed by Brown (1).[4] It consisted of seven stimulus lamps mounted in a horizontal row on a black plywood panel and spaced at 8-degree intervals. The panel was curved so that all lamps were 5 ft. from *S*, who was seated in front of a stand on which was mounted a spring-return toggle switch. When *S* moved the switch forward or backward, the words "win" or "lose" would light up in the upper or lower *left* quadrants of a small box below the central lamp. The upper and lower *right* quadrants also contained the words "win" and "lose," and were activated by *E* to give *S* information as to the actual outcome of each trial.

Procedure. The instructions included the following features: *S* was told that each of the seven lamps represented a race horse. When a lamp was lighted, it indicated that this horse was running a race against other horses not represented on the stimulus panel. The seven horses on the panel were not competing against each other. Each time a light went on, it was a new race, and *S* was supposed to guess whether the horse would win or lose that race. If *S* thought the horse would win, he was to push the switch forward. If he thought it would lose, he was to pull the switch toward him. *S* was told to remember how well each horse did in order to increase the accuracy of his bets.

Each trial (or race) consisted of the presentation of a stimulus light by *E*. *S*'s response automatically terminated the stimulus, at which time *E* activated the appropriate light to indicate the actual outcome of the race. Unknown to *S*, a predetermined schedule of 80 per cent win for the central lamp but only 20 per cent win for the other six lamps was used. There were 210 randomly ordered races. The response measure was the frequency of win responses to each stimulus lamp.

A Lindquist Type I trend analysis (3) was used to test the difference in the shape of the gradients elicited from the high and low MAS groups.

RESULTS AND CONCLUSIONS

The data for all of the psychiatric *S*s ($N = 169$) indicate that the frequency of win responses decreases progressively on each side of the central lamp as the angle separating a given lamp from

---

[1] A modification of this paper was read at the Midwestern Psychological Association meetings in Chicago, 1957.

[2] The assistance of Julia Weinberg and Jon Weinberg is gratefully acknowledged.

[3] Division of Psychology, Department of Psychiatry College of Medicine.

[4] The authors wish to express their appreciation to Judson S. Brown for his consultation and suggestions concerning the apparatus.

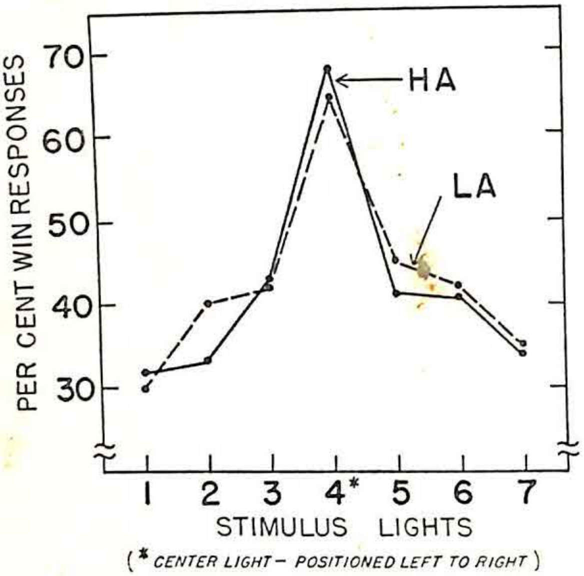

FIG. 1. PER CENT WIN RESPONSES FOR THE HIGH AND LOW ANXIETY GROUPS

the center increases. The difference in gradients among the six peripheral lights was significant beyond the .001 level. Presumably, the response tendency to bet "win" to the central stimulus generalized to the peripheral stimuli with those closest to the center most affected. This result with psychiatric *S*s is in keeping with Brown's findings with college students.

Figure 1 presents the percentage of win responses plotted for the high and low MAS groups. The results show no evidence of differences and corroborate Buss's findings.

In the light of the studies cited, these results suggest that there is no relationship between MAS and stimulus generalization in psychiatric *Ss*. Moreover, situational factors do not seem important in limiting the generality of such an interpretation. Since the relationships between anxiety and learning phenomena are generally well recognized, these negative results presumably reflect the inadequacy of the Taylor scale as a relevant index of anxiety levels in psychiatric subjects.

REFERENCES

1. BROWN, J. S., CLARKE, F. S., & STEIN, L. A new technique for studying spatial generalization with voluntary responses. *J. exp. Psychol.*, 1958, **55**, 359–362.
2. BUSS, A. H. Stimulus generalization as a function of clinical anxiety and direction of generalization. *J. abnorm. soc. Psychol.*, 1955, **50**, 271–273.
3. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
4. ROSENBAUM, G. Stimulus generalization as a function of clinical and experimentally induced anxiety. Unpublished doctoral dissertation, State Univer. of Iowa, 1950.
5. TAYLOR, JANET A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.
6. WENAR, C. Reaction time as a function of manifest anxiety and stimulus intensity. *J. abnorm. soc. Psychol.*, 1954, **49**, 335–340.

Received January 18, 1958. Prior publication.

THE ACCEPTANCE OF INFORMATION INTO ATTITUDE STRUCTURE[1]

HELEN PEAK AND H. WILLIAM MORRISON
University of Michigan

It has been demonstrated in a number of experiments that attitudes toward policies or actions may be changed by providing information[2] about their consequences (2, 5, 9). These studies have shown, however, that persons in positions at the opposite extreme from a communication are less likely than others to change attitudes, in spite of the fact that many of them are influenced by the experimental procedure. For example, in Carlson's experiment (2), which modified attitudes by this method, there was a change in information structure that was not always reflected in change of attitude. Thus, the selectivity that results from possessing certain attitudes does not necessarily prevent or distort the perception of information, nor does it necessarily influence what is remembered. But in these experiments it did have an effect on the assimilation of information into the structure of the attitude.

The opportunity to examine further the relation between attitude position and the acceptance of information into a structure presented itself in connection with an experiment designed to study the relation of certain personality characteristics and susceptibility to change pressure, in which we employed a modification of the Carlson technique to alter attitudes toward segregation.

METHOD

One hundred and sixty-nine students registered in the elementary psychology course at the University of Michigan completed all parts of the experiment reported here. Minority group members (Negroes, Jews, Southerners, and persons born abroad) were excluded from the sample since the segregation issue clearly has special meaning for these persons.

Data were collected during the fall of 1955. During the first and third sessions, which were regular class periods, measures were administered by the instructors as "illustrations of psychological procedures." Graduate student assistants met the group outside of class for observations at Time 2.

During Session 1 a number of tests were administered, which included among others the California F Scale (1) and scales for measuring attitudes toward six issues (9), including Negro housing segregation. Four to five weeks later in Session 2, change procedures were administered in experimental groups, while controls were given other tasks. Three weeks subsequently in Session 3, attitudes were again measured, and *Ss* were asked to list all the good consequences that people might suggest to be the result of removing Negro housing segregation. They were then instructed to underline those results that they accepted as their own. As the next task, they listed bad consequences and underlined those accepted. The order of listing good and bad consequences was counterbalanced from person to person.

Control and experimental groups received identical treatment at Times 1 and 3. At Time 2, controls performed tasks that were unrelated to the attitudes being measured. Procedures at Time 2 were the same for both experimental groups, except for the direction of pressure. The *Ss* were told that a test of objectivity was being devised and that one mark of this characteristic is the ability to present effective arguments on any side of an issue regardless of one's own point of view. One statement, about an alleged good consequence of removing segregation for the antisegregation group, or about an alleged bad consequence for the prosegregation group, was written on the top of each of four sheets of paper. For six minutes *Ss* listed reasons in support of one statement. They then turned to the next and did the same for another statement in support of the same view, and so on. A given group dealt only with good or bad consequences of removing segregation, not with both. After each of four statements was elaborated in this way, "scoring" sheets were handed out, listing a number of points that "might have been made in support of each statement." *Ss* were asked to mark in spaces provided on the scoring sheet, those consequences which corresponded with the ones that they had written; and in other spaces, to check the ones they had failed to give. It was hoped that by this device they would be induced to read and consider all the arguments set down on the side of the issue with which they were working. An attempt was made to parallel arguments in favor of segregation and those against it, so that the number of items of information provided in the communication would be the same for the two experimental groups. Within the Time 2 session, the same sort of task was carried out for attitudes toward the location of airports.[3]

---

[1] This study is part of a project carried out under Contract Nonr-1224(10) NR 171-039 between the University of Michigan and the Office of Naval Research. Mark Spivak and J. L. Zinnes helped in the collection and tabulation of the data.

[2] A unit of "information" is a statement or portion of a statement which argues for, or presents support for, or rationalizes some attitude position.

[3] These data have not been analyzed. Presumably there is less personal involvement in that issue, so that the determinants of behavior in that domain should be less closely related to the *F*-scale measure.

Attitude Measures

Three different scales measuring attitudes toward segregation were administered (9). These scales proved to have high intercorrelations ranging from +0.65 to +0.95. In the analysis reported here, a ranking scale developed from Coombs' unfolding technique was used to establish attitude position (2, 4, 11). On this scale, which had a test-retest reliability of +0.68, statements were ranked in terms of their acceptability, and scores were based on those statements assigned the first three ranks. The statements ranged from, "I am completely opposed to allowing Negroes in white neighborhoods" to "I am completely in favor of allowing Negroes in white neighborhoods." These and four intermediate statements (9) followed a brief introductory statement of the segregation problem.

Information Measures

The *S*s listed reasons for or against removing Negro housing segregation under the instructions already described. The material obtained in this way was coded in terms of basic items of information. For example, a single statement by *S* was scored each time mention was made of the existence of certain conditions, of attitudes held by various groups, of traits characteristic of a group, or of consequences that might follow from desegregation.[4] The method of scoring a complex statement for these more basic pieces of information may be illustrated by this example:

> Since Negroes are naturally sloppy (*Trait*), most Negro neighborhoods are dirty and unsanitary (*Present Condition*), and white people don't like to live in the same neighborhood (*Attitude*). Therefore, mixing the two groups would lead to discontent and violence (*Consequence*) and a lowering of property values in the neighborhood (*Consequence*).

Each of the units of information thus isolated was coded in order to make these distinctions: (*a*) whether or not the item was included in the change communication; (*b*) whether *S* regarded it as a good or a bad consequence or characteristic; and (*c*) whether *S* accepted the item or only knew about it. The third distinction was made on the basis of whether *S* had underlined a statement. The interpretation that an item implied a good or bad characteristic was determined by whether *S* had written the item on the sheet that called for good or for bad consequences. In order to determine whether or not items of information were or were not like points that had been made in the communication, a code was prepared which distinguished two degrees of similarity to communication items, and the analysis has been made on the basis of information units classified as being "in" or "out" of the communication. The coder had no knowledge of the attitude position of the individuals whose remarks he was scoring.

These protocols are the source of the material analyzed in the following pages. In the presentation, *information known* refers to items written down by an *S* as a characteristic or consequence of desegregation, without regard to whether or not he accepted the statement as true or relevant to the issue. *Information accepted* derives from the same source but includes only that material which the *S* indicated he could accept. From time to time reference will be made to the *acquisition of information*. Given the design of the present experiment, the inference that information has been learned or acquired by contact with the communication can be made only by a comparison between the number of items produced by experimental groups who saw a communication and controls who did not.

HYPOTHESES AND RESULTS

Attitude Position and Related Sources of Affect

In the light of evidence from earlier studies (2, 5, 10), it is predicted that attitudes toward an issue are positively correlated with the algebraic sum of the number of good and bad consequences accepted as characteristic of the object of the attitude.

Table 1 presents evidence of the relation between attitude position and the number of items of information accepted by persons in the control group. Whether we consider information that is in or out of the communication, prosegregation items, antisegregation items, or the difference in the number of pro- and antisegregation items, it is clear that the hypothesis is confirmed. Those who are against segregation have accepted more antisegregation arguments, while those at the other end of the attitude scale accept more prosegregation material. A one-way analysis of variance indicates that these differences are significant at the .05 level or less.

Attitude Position and Amount of Information Known

It might be thought that those who hold extreme attitudes would possess more information about the object of these feelings than would less involved persons, and that the observed smaller change of attitude in the former group was related to the lack of new information in the change communication. On the other hand, it is equally plausible to suppose that the highly informed person would take a moderate position because the *pros* and *cons* tend to cancel each other. These considerations led us to test the null hypothesis: the total amount of information known about the object

---

[4] Each item that was produced by *S* was counted even though he produced others similar to it. A partial analysis of nonredundant items showed the same relations of attitude position and number of good and bad consequences as we obtained by the redundant count reported here.

TABLE 1

MEAN NUMBER OF INFORMATION ITEMS ACCEPTED IN RELATION TO ATTITUDE POSITION

(Control group)

Items of Information		Attitude Position				F	p
		Anti-segre.			Pro-segre.		
		1	2	3	4		
Anti minus pro items in communication	M	+1.88	+2.40	+0.43	−1.23		
	σ^2	6.75	11.14	7.39	8.18	6.53	<.005
Anti minus pro items in & out	M	+2.80	+2.15	+0.29	−2.05		
	σ^2	11.04	12.43	10.92	12.50	8.56	<.001
Anti items in communication	M	3.80	4.15	2.93	2.32		
	σ^2	2.88	6.53	3.64	4.76	3.14	<.05
Pro items in communication	M	1.92	1.75	2.50	3.55		
	σ^2	3.19	3.79	4.25	5.97	3.16	<.05
Anti items in & out	M	5.28	5.00	4.29	2.96		
	σ^2	5.00	6.00	5.77	5.95	4.04	<.025
Pro items in & out	M	2.48	2.85	4.00	5.00		
	σ^2	6.01	5.53	8.00	8.36	4.03	<.025
	N	25	20	14	22	$df =$	3/77

TABLE 2

MEAN NUMBER OF ITEMS OF INFORMATION KNOWN IN RELATION TO ATTITUDE POSITION

(Control group)

Items of Information		Attitude Position				F	p
		Anti-segre.			Pro-segre.		
		1	2	3	4		
Anti plus pro items in & out	M	12.44	15.10	13.85	15.09		
	σ^2	23.53	36.59	35.84	21.99	1.8	ns*
Anti items in communication	M	4.76	5.70	4.64	4.64		
	σ^2	4.02	5.11	5.66	6.05	1.0	ns*
Pro items in communication	M	3.72	3.95	4.50	5.54		
	σ^2	3.96	5.75	5.96	3.70	3.02	<.05
Anti items in & out	M	6.98	7.90	6.79	6.32		
	σ^2	7.77	12.04	9.63	9.78	0.86	ns*
Pro items in & out	M	5.46	7.20	7.07	8.77		
	σ^2	9.00	14.11	12.82	8.83	3.76	<.025
	N	25	20	14	22	$df =$	3/77

* Not significant.

of an attitude, including items accepted and not accepted, does not differ as a function of attitude position.

The one-way analysis of variance of total information in Table 2 does not allow rejection of the hypothesis. The difference in total information in the four attitude categories is not significant. However, when the number of items of information at Position 1 is compared with the number at the opposite extreme, there is a significant difference with $t = 4.5$ and $p < .01$. Among those favoring segregation there is, then, somewhat more total information, although differences over all positions are not significant.

Table 2 also indicates the relation of attitude position to the mean number of antisegregation and prosegregation items. There is a significant relation between attitude position and the number of prosegregation items known, those who favor segregation producing more supporting information about it.[5] However, the same is not true of antisegregation information. Those who are in favor of desegregation do not have more information that supports their side of the issue than do persons at the opposite end of the scale.

The fact that persons in favor of segregation have somewhat more information about their side of the issue than do those opposed brings up the problem mentioned earlier. In presenting a communication for the purpose of changing attitudes, it is important to be sure that the amount of change produced is not restricted by availability of new information. In spite of the difference in prosegregation items known, it appears that in the present case this factor can be ruled out as an explanation of any differential reactions to the communication that are related to attitude position. In a comparison of the communication items actually presented with those known by *S*s there were obviously many points that could have been acquired from the communication even by those with maximum prosegregation knowledge.

Compatibility of Information, Attitude Position, and Communication Pressure

The availability of new information discussed above is not the only determinant of the effects of a communication. Granting that at times the perception and retention of material may be affected by the compatibility between incoming information and existing knowledge, we suspected that this factor of compatibility would be even more important in determining whether information was *accepted* as part of an attitude structure. For example, *S* may hear and remember a communication presenting arguments that segregation is bad, but such an item may simply remain part of a recollection that the statement was made, instead of being accepted as relevant to the problem of segregation, unless it is in some degree congruent with existing attitudes.

---

[5] M. B. Smith has suggested that the finding that prosegregation *S*s in the control group know more prosegregation information than antisegregation controls, might be based on the latter's unwillingness to dignify certain prosegregation statements by the term "reasons." This interpretation is consistent with the finding that following the communication experimental *S*s who hold antisegregation positions (having now learned that such statements are considered "reasons") know significantly more prosegregation information than comparable controls.

The most appropriate test of this hypothesis would involve observations of both attitude position and information before and after the presentation of a communication. Since the present study provides only post-change measures of information in control and experimental groups, it is only possible to compare these two groups and to attribute information differences in them to the effects of the communication. In order to investigate differences in the information of experimental and control *S*s who endorse a particular position, *S*s with similar Time 3 attitude positions are compared.[6]

The hypotheses are: 1. Those who have seen a communication supporting one side of an issue and who assume an attitude position compatible with that view, accept more information from such a communication than do controls in the same attitude position. 2. Those taking a position incompatible with the communication are not different from controls in items accepted in the direction of the communication (null hypothesis). 3. The amount of information known has less relation to the compatibility between communication and attitude than amount accepted.

Acceptance of Communication Information. Table 3 reveals that persons in favor of desegregation accept significantly more antisegregation items under antisegregation pressure than do controls, although they do not accept significantly more items under prosegregation pressure. In this case, therefore, *S*s whose attitudes were compatible with the communication accepted more information than *S*s whose attitude position was incompatible with the communication. This was predicted.

In the case of those favoring segregation and receiving a prosegregation communication, the prediction is not supported, for there is not a significant excess of accepted prosegregation items in the experimental group when compared to controls. Under antisegregation pressure, persons oriented in favor of segregation did not accept significantly more antisegregation items than controls, as was expected.

Thus, the first part of the hypothesis is partially confirmed, and there is no basis for rejecting the second (null) hypothesis.

Knowledge of Communication Information. Table 3 indicates that in the experimental group, which experienced antisegregation pressure, persons favoring desegregation know significantly more compatible communication items than do controls in the same attitude position, and the difference decreases the closer attitudes are to the prosegregation end of the scale. But the same result emerges under prosegregation pressure. Here, again, antisegregationists have more knowledge of the communication, while those favoring segregation do not differ significantly from controls. Indeed, the only significant difference between experimental and control persons with prosegregation attitudes occurs in those who saw the antisegregation communication; in this case, the experimentals know significantly *fewer* prosegregation arguments than do controls ($p = .04$ for items in and out of communication and 0.1 for those in communication). Those with the opposite view have no tendency to reduce the output of information under any of the conditions studied.

In short, the amount *known* from a communication is not closely related to the compatibility between attitude position and the communication point of view. This was predicted. The amount known is more definitely related to attitude position, and those opposed to segregation have more information about the communication material than those in favor of segregation, regardless of the direction of the pressure.[7]

---

[6] It has seemed more reasonable to break the data on Time 3 than on Time 1 attitude position, although a partial analysis of the information patterns that appear with a Time 1 break do not appear to differ greatly from those presented here. There was not a great amount of attitude change as a result of the communication.

[7] See Footnote 5.

F Scores, Information, and Attitude Position

Even though the preceding hypothesis predicts a general tendency to make attitudes and accepted statements compatible, it must also be assumed that individuals differ in the extent to which dissimilar information is permitted to exist within any structure. For example, it has been suggested that those with authoritarian personality characteristics require more similarity among the traits which they accept as belonging together than do others (1, 6, 11). Klein's concepts of levellers and sharpeners also deal with individual differences in the degree of similarity demanded before events are reacted to as the same or different (8).

It was expected that high Fs might have rejected more information (both known and accepted) before hearing the communication. There was no reason to believe, however, that low Fs would necessarily accept more new information from the communication than high Fs, although we did expect them to remember more of this material since they would presumably have less tendency to ignore what they could not assimilate. Therefore, it is predicted: 1. In the control group, individuals with high F scores know and accept fewer items of information incompatible with their attitudes than do low Fs at the same attitude position. 2. Following change pressure, low Fs know more items of information incompatible with their attitudes than high Fs, when each is compared with its control group.

Analysis of Control Group. Although there is no significant difference between high and low Fs in the amount of information known or accepted, further analysis makes it clear that F scores are related to information patterns. Since there is not a significant correlation between F scores and attitude position in this study (9, p. 20), it has been possible to obtain a sample of low Fs and high Fs from both halves of the distribution of Time 3 attitude positions. An analysis of the sources of variance in information material when broken by F scores and attitude position suggests that the major contribution of the authoritarian syndrome is to be found in its interaction with attitude position. This interaction is significantly related ($p < .05$) to antisegregation items accepted, to prosegregation items accepted, to the sum of both accepted, and to the combination of pro- and antisegregation information known.

TABLE 3

EXPERIMENTAL MINUS CONTROL GROUP DIFFERENCES IN MEAN NUMBER OF COMMUNICATION ITEMS ACCEPTED OR KNOWN IN THE DIRECTION OF COMMUNICATION

Direction of Pressure and Kinds of Items		Attitude Position			
		Anti-segre.			Pro-segre.
		1	2	3	4
Accepted					
Prosegre. pressure, Prosegre. items	Diff. Means	−0.21	+0.58	−0.21	+0.30
	p*	ns	ns	ns	ns
Antisegre. pressure, Antisegre. items	Diff. Means	+1.99	+0.15	−0.33	−0.32
	p*	<.01	ns	ns	ns
Known					
Prosegre. pressure, Prosegre. items	Diff. Means	+2.42	+2.05	+1.70	−0.60
	p*	<.01	<.04	<.03	ns
Antisegre. pressure, Antisegre. items	Diff. Means	+1.95	+1.60	+0.66	+0.79
	p*	<.015	<.07	ns	ns
N control		25	20	14	22
N pro-pressure		7	6	14	20
N anti-pressure		14	10	10	7

* Based on t test.

Some light is shed on the meaning of the interaction between F scores and attitude position by noting the nature of the variation in information when either attitude position or F score is held constant. Table 4 indicates the average number of information items (in and out of communication)[8] in each category and the significance of the differences between means. When differences between those persons holding different attitudes but having similar F scores are examined, it appears: (*a*) When Ss have low F scores and are opposed to segregation, they know and accept more items supporting this view than do those low Fs who favor segregation. This fact gains interest when set over against the next point. (*b*) When Ss have high F scores and are opposed to segregation, they do not differ in the number of antisegregation items known or accepted, from high Fs who take the opposite view.

---

[8] An analysis based on items in the communication produced essentially the same results.

However, these two groups of people do differ from each other in the number of items favoring segregation. In other words, attitude position is correlated with amount of information supporting segregation in high *F*s and with amount of information against segregation in low *F*s.

These results suggest differential sensitivity of persons with high and low *F* scores to the two kinds of information. The material in favor of segregation tends in many cases to imply distrust of people, and it is this kind of information that differentiates the high *F*s with different attitudes. On the other hand, material opposed to segregation, which suggests good qualities in the minority group, differentiates low *F*s (1). In this case, our findings might derive from either of two sources. High *F*s who are against segregation may simply lack prosegregation information which, if available, would be readily integrated into their attitude structure to produce a shift in attitude; and vice versa for low *F*s. But this seems an unlikely explanation because of the general availability of both kinds of information to people in a college population.

The more reasonable suggestion would appear to be that the small amount of prosegregation information produced by the high *F* who is against segregation and the small amount of antisegregation information produced by the low *F* who favors segregation results somehow from a neglect or denial of that kind of material which theoretically should be of most interest to these people. This interpretation implies that there is some sort of conflicting pressure for these people, operating to prevent acceptance of such information. It might be reasonable to suppose that in the present climate of opinion, attitudes toward the segregation issue are very crucial for many people, with the consequence that if a social group takes a stand it is likely to be a strong stand, whatever its direction. The result for some individuals may be a tendency to maintain attitudes enforced by an important social group, even against the kind of personal predilections represented by *F* scores.

A related analysis deals with the variation in the information pattern of those with similar attitudes but different *F* scores (Table 4). Low *F*s who are against segregation know and accept more total information than do high *F*s with similar attitudes. This difference is significant only for total items accepted, though it is in the same direction for both pro and antisegregation items accepted or known. On the other hand, low *F*s who are in favor of segregation know fewer items of information (pro- or pro-plus antisegregation) than do high *F*s with similar attitudes. Again the differences are all in the same direction, though only those mentioned are significant.

Once more these data suggest a tendency for certain groups of individuals to neglect information. If incompatibility between *F* structure and attitude position implies conflicting pressures, then we would expect this observed reduction in the total amount of information known or accepted by people with such incompatibility. Not only do certain kinds of information appear to be repressed, but there is a general reduction of output by those with incompatible structures. Combining total amount of information accepted or known by high *F* prosegregationists and low *F* antisegregationists, we find significantly more information output ($p < .05$) in this group with compatible structures than in the group made up of high *F* antisegregationists and low *F* prosegregationists (the "incompatibles").

Experimental and Control Groups Compared. It was predicted that following a communication, low *F*s would know more information incompatible with their attitudes than would high *F*s in comparable attitude positions. The results indicate that neither high nor low *F* individuals with prosegregation attitudes know significantly more antisegregation material following antisegregation pressure than do controls. The same is true of both high and low *F*s with antisegregation views after they had seen a prosegregation communication, when compared with controls on number of pro items known. None of these groups acquired a significant amount of incompatible information, and low *F*s did not differ from high's in this respect.

But once again these negative results are modified by a more detailed analysis of the material. There are differences in high and low *F*s, though not precisely those predicted. Table 5 summarizes the significant differences in experimental and control information (in and out of the communication)[9] and suggests these conclusions:

---

[9] See Footnote 8.

TABLE 4

MEAN NUMBER OF ITEMS OF INFORMATION ACCEPTED OR KNOWN, IN RELATION TO ATTITUDE POSITION AND F SCORE

(Control group, in and out of communication)

	Number of Pro Items			Number of Anti Items			Number of Anti + Pro Items		
	Antisegre. Attitude**		Prosegre. Attitude**	Antisegre. Attitude		Prosegre. Attitude	Antisegre. Attitude		Prosegre. Attitude
Items Accepted									
Low F	3.08	(ns)*	4.00	5.87	(<.01)	3.25	8.96	(ns)	7.35
	(ns)		(ns)	(ns)		(ns)	(<.05)		(ns)
High F	2.14	(<.015)	5.08	4.33	(ns)	4.14	6.48	(<.05)	9.19
Items Known									
Low F	6.42	(ns)	7.50	7.74	(<.05)	5.70	14.16	(ns)	13.20
	(ns)		(ns)	(ns)		(<.05)	(ns)		(<.05)
High F	6.02	(<.01)	9.06	6.98	(ns)	7.74	13.00	(<.05)	16.80

N (Pro-Low F) = 20 N (Pro-High F) = 21
N (Anti-Low F) = 24 N (Anti-High F) = 21

---

* Items in parentheses are probabilities based on t test that differences in means are due to chance; (ns) means not significant.

** Attitude positions are broken into two categories here instead of the four previously used on account of the limitation in number of cases.

TABLE 5

EXPERIMENTAL MINUS CONTROL GROUP DIFFERENCES IN MEAN NUMBER OF ITEMS ACCEPTED OR KNOWN

(In and out of communication)

Pressure Direction		Prosegregation Attitudes						Antisegregation Attitudes					
		Accept Pro	Accept Anti	Accept Pro & Anti	Know Pro	Know Anti	Know Pro & Anti	Accept Pro	Accept Anti	Accept Pro & Anti	Know Pro	Know Anti	Know Pro & Anti
Prosegregation Pressure	Low F Diff. Mean	+1.11	+0.59	+1.71	+1.66	+2.06	+3.80	+0.04	−2.00	−1.96	+2.58	+1.50	+4.08
	p*	ns N = 38	ns	ns	ns	<.05	<.05	ns N = 32	ns	ns	ns	ns	ns
	High F Diff. Mean	−1.17	−1.95	−3.13	−1.96	−1.36	−3.30	−0.74	−0.13	−0.88	+0.88	+0.72	+1.60
	p*	ns N = 30	<.05	<.05	ns	ns	ns	ns N = 26	ns	ns	ns	ns	ns
Antisegregation Pressure	Low F Diff. Mean	−0.80	−1.15	−1.95	−1.60	+0.80	−0.80	−0.01	−0.41	−0.42	+1.16	+0.60	+1.76
	p*	ns N = 30	ns	ns	ns	ns	ns	ns N = 37	ns	ns	ns	ns	ns
	High F Diff. Mean	−2.62	+0.57	−3.19	−3.14	+0.18	−3.19	+0.04	+2.94	+2.98	+1.38	+2.98	+4.36
	p*	<.05 N = 28	ns	<.05	<.01	ns	<.05	ns N = 32	<.01	<.05	ns	<.05	<.05

* p based on t test.

a. High Fs typically show either an excess of information in the direction of the communication pressure or a deficit of information opposed to the communication, which in either case appears to reflect the familiar tendency of high Fs to conform (1, 7, 3, p. 134). But there is one exception to this conforming behavior of those with high F scores. Those who have antisegregation attitudes following the prosegregation communication show neither more nor less information than comparable controls. This finding is of interest when it is recalled that in the control group analysis it was the high Fs in this same attitude position who knew and accepted the smallest amount of information. We return to this point below.

b. When low F experimentals differ from their controls, they know more anti- *and* more prosegregation information than the controls. In this they are unlike high Fs who depart from their controls only in the direction of greater conformity with the communication. Moreover, only the differences in items known are significant in the low F group, whereas the high Fs sometimes exceed controls in the number accepted as well as the number known. This finding again suggests greater conformity on the part of the latter.

c. A final observation reverts to the "incompatibles" mentioned earlier, the low F with prosegregation attitudes and the high F who opposes segregation. These two groups, following a communication inconsistent with their attitude position, and presumably consistent with their F scores, show no difference from controls in items known or accepted. It can only be surmised that they resist the pressure incompatible with their attitudes for the same reasons that they produce less information than the "compatibles" in the control group analysis. If this is a correct analysis of the factors involved, it is apparent that *both* high and low Fs are disposed to defend against information incompatible with an adopted attitude, which is itself in conflict with some aspect of the personality organization.

ATTITUDE CHANGE AND ACQUISITION OF INFORMATION

The design of the experiment does not make possible a direct correlation of attitude change with change in information because, as already noted, the information measures were made only at Time 3. But it would be expected from other experiments (2) that such a relation should exist, and the present results tend to support such a hypothesis. It has been reported elsewhere (9) that in this experiment significant attitude change in the direction of the communication was produced only under antisegregation pressure (p difference experimental and control $= .05$). Similarly, we have just seen that when the groups are taken as a whole, the only significant difference between experimental and control groups in the amount of communication information *accepted* occurs in the case of those in the antisegregation position who experienced antisegregation pressure. These Ss accept more antisegregation items than do comparable controls, while under this same pressure, persons who assume prosegregation attitudes accept fewer items of prosegregation information. In other words, although they do not indicate acceptance of more communication items than controls, they do accept a smaller amount of information opposed to the communication. These are the results to be expected if acceptance of information is correlated with attitude change.

SUMMARY

Attitudes toward Negro housing segregation were determined before and after providing college students in an experimental group with arguments about segregation, in a manner intended to produce involvement in the procedure. In a control group, attitude measures preceded and followed a session devoted to tasks unrelated to segregation. Finally, all Ss listed good and bad consequences that people might suggest to be the result of removing segregation. These protocols yielded evidence about the nature and amount of information *known* and *accepted* under various conditions. "Information" refers to the content of arguments relevant to the issue.

1. In the control group, attitude position is significantly related to pro- and antisegregation items of information *accepted* by Ss (predicted).

2. As expected, in the control group amount of information *known* does not consistently vary with attitude position. There is some tendency, however, for prosegregation Ss to *know* more pro- than antisegregation items (not predicted).

3. The hypothesis that Ss who have seen a communication supporting one side of an issue and have assumed an attitude position compatible with that view, have more *accepted* information than comparable controls, is partially confirmed.

4. As expected, the difference between experimental and control Ss on amount of information *known* is not closely related to the compatibility of attitude position and communication point of view. The significant differences are closely related to attitude position.

5. A significant interaction between attitude position, F-scale score, and kind of information

(*pro* or *anti*) determines information *known* and *accepted* in the control group. Total information produced is related to the interaction of F score and attitude position.

6. Significant differences between F groups receiving the communication and comparable controls indicate that the high Fs tend to conform more to the communication, while low Fs tend to *know* more information both congruent with and opposed to the communication position. Cases where the communication produced no effect were noted.

The immediately preceding results (5 and 6) were more complex than predicted. They were interpreted in terms of the degree of compatibility of F syndrome predispositions and adopted attitude position, which under certain pressures result in accepting patterns of information incompatible with personal predilections.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWICK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. CARLSON, E. R. Attitude change through modification of attitude structure. *J. abnorm. soc. Psychol.*, 1956, **52,** 256–261.
3. CHRISTIE, R., & JAHODA, M. *Studies in the scope and method of "The authoritarian personality."* Glencoe, Ill.: Free Press, 1954.
4. COOMBS, C. H. A theory of psychological scaling. *Univer. of Michigan Engineering Research Bull. No. 34*, Ann Arbor: Univer. of Michigan Press, 1952.
5. FELD, S. The persistence of attitude change. Unpublished study, Univer. of Michigan, 1953.
6. FRENKEL-BRUNSWICK, E. Personality theory and perception. In R. R. Blake & G. V. Ramsey (Eds), *Perception.* New York: Ronald, 1951.
7. FROMM, E. *Escape from freedom.* New York: Rinehart, 1941.
8. KLEIN, G. S. The personal world through perception. In R. R. Blake & G. V. Ramsey (Eds), *Perception.* New York: Ronald, 1951.
9. PEAK, H. *et al.* Some factors in resistance to attitude change. I. Personality factors. Technical Report, No. 1, Office of Naval Research, August 15, 1956.
10. ROSENBERG, M. J. Cognitive structure and attitudinal effect. *J. abnorm. soc. Psychol.*, 1956, **53,** 367–372.
11. STEINER, I. D. Ethnocentrism and tolerance of trait "inconsistency." *J. abnorm. soc. Psychol.*, 1954, **49,** 340–354.

Received July 22, 1957.

PERSONALITY VARIABLES AND RECOGNITION RESPONSE LEVEL[1]

ARNOLD BINDER
Indiana University

Visual recognition may be conceived of as a process of categorization in which an available set of object attributes (cues) are used to assign a given object to a class. An assigned response class or category may be a name like "orange" if the object is spherical, orange-colored, about three inches in diameter, etc., or it may simply be the push of a right- rather than a left-hand lever. Thus, Binder (1) and Bruner (3) have hypothesized that a person in a perceptual recognition situation gathers cues for purposes of discrimination and responds with a class name when he has enough cues to satisfy himself as to the object's identity. The information in the form of cues serves to lower the observer's uncertainty as to the class name of the object; that is, an increasing number of available cues leads to a progressively more restricted number of possible class assignments.

Hilgard (8), Miller (10), and Binder (1) have hypothesized that people tend to give a class-naming response before they have enough cue information for unique specification or, in other words, that people leap to inferences on the basis of inadequate cue data. The purpose of the present research was to investigate the tendency to respond early in the process of cue accumulation, before unambiguous specification is possible, as a function of personality characteristics and/or abnormalities. Smock's investigation (15) of the tendency to respond to a stimulus configuration when insufficient discriminative information is available is most closely related to this research. He found that subjects (*S*s) under "stress" conditions respond earlier (at levels of greater uncertainty) than *S*s under "security" conditions and concluded, ". . . anxious individuals tend to resolve ambiguous or unstable situations through premature structuring and closure" (p. 181).

A number of recent studies involving measures of reaction speed rather than point of response in a sequence of increasing stimulus information are also relevant and most useful in generating experimental hypotheses. Thus, Eriksen (5) defined "hysterics" on the basis of high values on the distribution of hysteria minus psychasthenia (*Hy* − *Pt*) Minnesota Multiphasic Personality Inventory (MMPI) scale scores and "psychasthenics" on the basis of low values on this distribution, and found that the hysterics had shorter median response latencies than the psychasthenics in a simple recognition task involving stress. And Grice (7) and Rockett (12) have shown that high anxious *S*s, as defined by the Taylor and Sarason Anxiety scales, respectively, have slower recognition reaction speeds than low anxious groups.

In terms of the analysis of perception described above, it is reasonable to expect people who respond in a recognition situation on the basis of little cue information to be more likely to have fast reactions. Similarly, those who respond only after gathering a good deal of information should have relatively slow reactions.

With these assumptions we can use Eriksen's results (5) to hypothesize that hysterics respond at higher levels of uncertainty than psychasthenics and the results of Grice (7) and Rockett (12) to hypothesize that low anxious *S*s respond at higher uncertainty levels than high anxious *S*s (when the various groups are appropriately defined). But since the very high relationship between Taylor and *Pt* scores has been convincingly demonstrated (2, 4, 6) and the evidence indicates that the Taylor and Sarason scales measure similar processes (9, 13, 14), we can use the Grice and Rockett data to hypothesize that low *Pt S*s respond at higher uncertainty levels than high *Pt S*s. The hypothesis based upon Eriksen's results, to be sure, can be expected to hold only when some degree of anxiety is present, but this restriction becomes less important for our purposes when consideration is given to the ample evidence that the conditions of most psychological experimentation are in themselves anxiety provoking (see Taylor [16] for a discussion of this point).

A final hypothesis stems from the finding of a pilot study conducted by the present writer.

---

[1] This research was supported by a grant from the Graduate School of Indiana University.

This is to the effect that higher paranoia (*Pa*) MMPI scale scores are associated with a tendency to withhold perceptual responses until a relatively low amount of uncertainty is present in the stimulus situation.

In summary, the purpose of the present investigation was, in general, to investigate the relationship between personality characteristics and amount of cue accumulation prior to recognition response and, more specifically, to test the following three hypotheses: (*a*) people with high *Hy−Pt* scores respond at higher uncertainty levels than low *Hy−Pt* scorers, (*b*) low *Pt* scorers respond at higher uncertainty levels than high *Pt* scorers, and (*c*) low *Pa* scorers respond at higher levels of uncertainty than high *Pa* scorers.

METHOD

Subjects

The experiment employed 35 *S*s, but the data from two of these could not be used because of administrative errors. The *S*s were all male students enrolled in two of the introductory psychology sections at Indiana University who volunteered by signing sheets circulated in the classrooms. The students were given the group form of the MMPI during regular class meeting hours.

Apparatus

From the viewing position of the *S*s, the exposure apparatus consisted of a 3′ by 4′ plywood panel mounted on a table. Roughly centrally placed in the panel were two 8¼″ by 8″ one-way vision mirrors, separated by an opaque strip 4″ wide and 8″ high. Attached to this panel, on the experimenter's (*E*'s) side, was a box containing solid bottom, sides, and back. The front part was open and the box was so attached to the panel that this front came directly behind the one-way mirrors. The top of the box was slotted, and the inner part of the sides and a central partition were fitted with grooves such that the exposure of the stimulus materials could be manipulated by the *E*. The stimuli were illuminated by four (for each mirror) six-watt bulbs which were placed behind the mirrors so as to provide relatively even stimulus surface illumination. Panel, table, and box were all painted a dull black.

The apparatus was situated in a semi-darkened room, and *S*s could see the stimulus materials only when the bulbs behind the mirrors were on. At other times, of course, they could see only the reflecting surfaces of the mirrors.

The timing of the on-off sequences of stimulus illumination was controlled by a bank of microswitches operated off cams turned by a constant speed motor. Since the experiment consisted of two phases, learning trials first and then test trials, two separate timing sequences could be produced by the proper manipulation of switches. For the learning phase, the left mirror lights (from *S*s' position) contained a stimulus object and went on about one and a half seconds after the sounding of an indicator buzzer. After these had been on for about three and a half seconds, the right mirror lights went on, making the proper response (name) for the stimulus object visible to *S*s. The two sets of lights remained on for about two seconds and then went off simultaneously. This condition prevailed for about nine seconds; that is, until about one and a half seconds after the buzzer indicated the start of a new sequence.

For the test phase of the experiment, only the left mirror was used. Again the lights behind it went on about one and a half seconds after the indicator buzzer sounded, but this time the stimulus figure remained visible for only about one and a quarter seconds. About a nine-second delay between the instant the lights went out and the next exposure enabled the *E* to change the stimulus cards.

Stimulus Materials

Two sets of stimulus cards were used in the experiment proper, one for the paired associates learning phase and one for the recognition test phase. The set used during learning, together with the associated responses or names, may be seen in Fig. 1. The test figures are shown in Fig. 2. In addition to these two sets, a third set was used for demonstration purposes immediately prior to the test phase. The first demonstration figure contained the two lines forming a caret at the bottom of the "lamppost" (see Fig. 1), the second demonstration figure contained these two lines plus the straight vertical line above it in the "lamppost," and the final demonstration figure contained all parts of the "lamppost" except the full circle at its top.

The stimulus figures were drawn by a draftsman, reproduced photographically, and then printed on dull photographic paper. These prints were heat-mounted on larger sheets of cardboard. The cardboard sheets were 12½″ wide by 11″ high, while the centrally mounted photographic paper was 11″ by 8½″. The figures were so arranged on the sheets that they appeared in the center of the left mirror (from *S*s' viewpoint) when the cards were placed in the appropriate exposure slots.

The test figures of less than full cues and the demonstration figures were obtained from the learning figures by the process of photographic masking. Great care was taken during original drafting, during photographic reproduction, and especially in the mounting process to be sure that cues common to two or more figures appeared in the same location of the exposure surface when the respective figures were placed in the slots.

The stimulus figure names were lettered on white drawing board, 12½″ wide and 11″ high. The placement of names on cards was arranged so that the names would appear in approximately the center of the right (from *S*s' viewpoint) mirror when the cards were placed in the slots behind it.

The learning set of stimuli consisted of nine objects, one introduced for the sake of facilitating test instructions (the "lamppost"), and eight experimental objects. These latter were designed such that all eight had common elements which, if presented to *S*s, provided no discriminative information (except to rule out the "lamppost"). Additional stimulus elements (cues) designated, successively, an object as belonging to one or another of groups consisting of four and two objects.

presented to *S*s sequentially so that each subsequent figure restricted the range of possible responses more than any of those preceding it. Using the labels for the stimuli of Fig. 2, a sequence of presentations might be A_1 (for one and a quarter seconds), lights out (nine seconds), B_1 (one and a quarter seconds), lights out (nine seconds), and so forth through C_2, D_4, and E_7. The sequence of stimulus presentations (in the left mirror) was stopped just as soon as a recognition response was made by *S*. The demonstration figures were used during the instructions preceding this phase but were not used subsequently.

In the instructions prior to the test phase, the *S*s were told that on each of several trials they would be shown a succession of figures in the left mirror for shorter periods of time than the objects shown during the learning phase. They were further told that no names would appear in the right mirror and that each succeeding figure in a given trial would contain more of the parts of one of the objects than the figure preceding it. It was emphasized that only one of the objects for which they learned the names would be involved in a trial of greater and greater cue completeness and that they were to indicate their recognition by giving the name of the object being represented just as soon as the recognition was achieved. In order to remove any possible lingering confusion on the part of the *S*s, the sequence of demonstration cards was then shown with the following instructions:

> Let us take an example: Suppose I select the object which is named "lamppost." Now I will present a series of cards in which the lamppost is more and more completely depicted.
>
> [As the demonstration figures were actually being shown in the left mirror of the apparatus.] Here is the first stage, if you recognize the lamppost at this presentation stage, you respond with the name immediately—before the next card is exposed.
>
> Now here is the second stage—you might recognize the lamppost for the first time here and respond accordingly.
>
> And the third stage where recognition and response might occur.
>
> Remember that each card will be presented for only a brief exposure and that a card of more cue information will follow it after a very short interval. It is important that you respond immediately upon recognizing the object represented.
>
> I will notify you when each new trial is to begin. At each new trial a different object will be represented in the sequence.
>
> Do you have any questions before we start the first experimental trial?

Each *S* was then presented with four sequences of figures of increasing cue content in the test trials. Stimuli A_1 and B_1 (see Fig. 2) always came first, in that order, and these were followed by either stimulus figure C_1 or figure C_2. Of the four test trials, two started with the sequence A_1, B_1, C_1 and two with A_1, B_1, C_2, but the order of their appearance in the test trial series was randomized for different *S*s. Next in the sequence, for each run, came either D_1 or D_2 if C_1 was the third figure, and either D_3 or D_4 if C_2 was the third stimulus. All of these fourth stage possibilities occurred once for each four test-run series, their order varying randomly. And, finally, the last stage for each run presented one of the two objects from Level E consistent in their cue composition with the immediately preceding stimulus, the choice between the two possible objects being random.

RESULTS

Each *S* was given a score of 1, 2, 3, 4, or 5 on each of the four trials, depending upon whether he gave a recognition response after the first, second, third, fourth, or fifth cue set, respectively. His recognition score was obtained by summing these four subscores. The maximum score of 20, thus, indicates that the *S* responded only after he had all cues necessary for complete discrimination available on all trials, while a score of 4 indicates that the *S* responded on all four trials when the available stimulus cue was part of all possible stimulus objects. The range of obtained total scores was 4 to 18, with distribution mean 11.5 and standard deviation 4.1.

The product-moment correlation between recognition score and each of the MMPI scale scores (with K correction) for the 33 *S*s may be seen in Table 1. Of the three scale scores (*Hy*, *Pa*, *Pt*) predicted to show some relationship with point of recognition response, only *Pa* showed a significant relationship. The correlation of .49 is significant beyond the .01 level.

For more sensitive tests of our three working hypotheses, it was desirable to compare the mean recognition score of *S*s with high MMPI scale scores with the mean recognition score of *S*s with low MMPI scale scores. Accordingly, all *S*s were ranked on the basis of their *Pt*, *Hy*, *Pa*, and $Hy-Pt$ score distributions separately and the mean recognition score of those falling in roughly the upper 50% of each distribution was compared with the mean recognition score of those falling in roughly the lower 50%. Because of the varying frequencies associated with each of the scores for the different distributions, it was not always possible to divide the entire set of *S*s into groups containing precisely 50% each. The rule accepted was to apply the cutting point between upper and lower half of the ranked MMPI score distribution in question to come as close to a 50–50 division as possible. Table 2 contains the *t* resulting from the comparison of mean recognition scores for groups falling at the upper and lower "half" of each of the MMPI score dis-

TABLE 1
PRODUCT-MOMENT CORRELATIONS BETWEEN RECOGNITION SCORES AND MMPI SCALE SCORES

F	*Hs*	*D*	*Hy*	*Pd*	*Pa*	*Pt*	*Sc*	*Ma*	*Si*
.34	.16	.22	.15	.40	.49	.16	.38	−.01	.20

TABLE 2
VALUES OF *t* IN THE COMPARISON OF MEAN RECOGNITION SCORES FOR *S*S DIVIDED INTO UPPER AND LOWER GROUPS ON BASIS OF MMPI SCORES

Score Distribution	Number of *S*s in Group		Mean Recognition Score		*t*
	Lower "Half"	Upper "Half"	Lower "Half"	Upper "Half"	
Hy	18	15	11.0	12.2	.79
Pt	17	16	11.4	11.6	.18
Pa	18	15	10.0	13.4	2.43*
Hy − *Pt*	15	18	11.5	11.5	.00

* $P < .05$.

tributions listed above. The only significant *t* was that resulting from the comparison of the mean recognition score of individuals who fell at the upper end of the MMPI *Pa* distribution with mean recognition score of individuals who fell at the lower end of this distribution.

Since the paired associates learning proceeded to criterion in this experiment, the *S*s of necessity had differing amounts of actual contact with the full cue stimulus objects. The question arises as to the relationship between these differences and the variations among recognition scores. To answer this question, recognition scores were correlated with number of learning blocks necessary to reach criterion and a product-moment coefficient of .24 was obtained. This value is not significantly different from zero. The correlations between number of learning blocks to criterion and MMPI scale scores vary between −.20 and .21, and are all insignificant. It would seem safe, then, to assume that the uncontrolled variable of actual number of learning trial blocks is not important for our present purposes.

In his experiment on recognition response uncertainty, Smock (15) found a significant difference between first response level during the early test trials and first response level during the later trials. Experience with the particular task apparently led to recognition responses earlier in the sequence of cue availability (at higher uncertainty levels). The mean recognition values over the 33 *S*s in the present experiment for the first, second, third, and fourth test trials were 2.64, 2.64, 3.09, and 3.15, respectively. There would seem to be a trend toward using more cues in later trials, rather than less as found by Smock. But a comparison of response levels during the first test trial with levels during the last trial by means of the sign test showed the difference between these levels to be nonsignificant. The test is not very powerful under the given conditions, but the limited range of possible values would make it unlikely that the assumptions for a more powerful trend test could be satisfied. In any event, the data for the present situation do not show the relationship reported by Smock.

One last result of some interest should be mentioned. At test levels B, C, D, and E (see Fig. 2), it was possible for the *S*s to give responses which were incorrect in being the names of objects which did not contain the available cues in their compositions. Thus, an incorrect response to figure C_2 would be "golf flag," to D_3 "banner," and to E_3 "vacuum cleaner." It was of course not possible to follow instructions and give an incorrect response to A_1. Out of a total of 20 Level B responses, 33 Level C responses, 40 Level D responses, and 11 Level E responses, the *S*s gave only one incorrect name. That was "lamppost" in response to B_1.

DISCUSSION

The obtained data confirm the hypothesis of relationship between *Pa* score and recognition score. The significant positive correlation indicates that people who get higher *Pa* scores tend to wait until a stimulus object is relatively clearly differentiated by a set of available cues before giving a naming response. This finding is surely consistent with the sensitivity, distrustfulness, and suspiciousness commonly found in persons who answer MMPI items in such a way as to give them elevated *Pa* scores. It would seem reasonable to conclude that the general distrust and suspiciousness of *S*s high in *Pa* led to a watch-and-wait attitude in the recognition task. They responded only when the available cues left little doubt or uncertainty as to an identifying name. It would seem worthwhile to test the deduction from this finding that high *Pa* scorers have significantly slower reaction speeds than low *Pa* scorers in tasks like those used by Eriksen (5), Grice (7), and Rockett (12).

The hypotheses stemming from the research in recognition reaction speed (5, 7, 12) on the relationship between *Hy* and *Pt* score, on the one hand, and recognition score, on the other, were not confirmed. In reaction speed experiments, the variable investigated, of course, is the inverse of delay between presentation of stimuli to *S*s and instant of response, with delay measured by a chronometer started simultaneously with stimulus exposure and stopped by voice key or some other response-triggered device. The tendency to accumulate more cues before responding is only one of many factors that could lead to decreased speed of recognition reaction. Differential speed in the actual process of cue accumulation and different numbers of competing response classes at a given point of cue accumulation are other factors that may account for the differences found among MMPI score groups in speed of recognition reaction. The evidence from this study indicates that there is little or no relationship between either *Hy* or *Pt* score (or their difference) and the tendency to require many cues before giving a naming response.

The formulations of Moffitt and Stagner (11) are of some interest in this context. These investigators administered five perceptual tests to *S*s differentiated on the basis of high and low Taylor Anxiety scale scores as well as on high and low situational anxiety (threatening vs. nonthreatening instruction groups). On the basis of their findings, using an interpretation of perception in terms of cue "scanning" and assimilation, they concluded that anxiety results in a diminished sampling of the cues necessary for making object discriminations. Anxious *S*s are thus handicapped in their ability to make the finer discriminations between overlapping cue sets. Referring to an anxious *S* in one example they state, " . . . if *S* samples a broken circle or other incomplete figure hastily, it is more likely that he will fail to sample the 'gap' and consequently will perceive the figure as complete . . . " (p. 356) failing, thus, to differentiate between a circle and a broken circle.

Using the reported high correlations between *Pt* score and Taylor score, the evidence from the present research does not support the hypotheses of Moffitt and Stagner, at least for the anxiety groups defined on the basis of Taylor score. It would seem that their formulations would involve a negative correlation between *Pt* and recognition score and, while the obtained correlation was not significantly different from zero, it was positive.

The correlation of .40 found between *Pd* and recognition score and that of .38 between *Sc* and recognition score are significant beyond the .05 level. The decision as to whether these obtained relationships are purely chance phenomena or representative of true relationships will have to await cross-validation. Surely the pattern seems to indicate that *S*s who answer MMPI items in the direction of maximum disturbance are least likely to make inductive inferences in a recognition situation on the basis of inadequate information. To paraphrase Miller (10), they do not tend to jump to conclusions from inadequate premises. Perhaps this has resulted from their having learned to distrust their environment as a result of many traumatic experiences with it.

A most interesting incidental finding in this investigation was the general lack of insight of the *S*s as to cue composition of the stimulus objects. This was very dramatically demonstrated by the efforts shown by many *S*s during the test trials to give the "right answer" even when responding at Levels A, B, C, or D (see Fig. 2). Obviously there is no one correct answer except for Level E responses, yet *S* after *S* asked if he had given the "right answer" to such test figures as A_1, B_1, C_2, and D_4. Some *S*s changed their responses from one to another, with signs of assurance at a clever move, when in fact the cues available at the cue stage were common to both figures. Common questions after the test trials were, "Was that right?" "Did I give the right answers?" and "Should I have said —— rather than —— to that last one?" All this despite the fact that only one incorrect test response was given in the entire experiment.

SUMMARY

This experiment was planned in the frame of reference of recognition as a process of classification or categorization. Recognition, in this context, is assumed to consist of assignment of a given stimulus figure to an appro-

priate class on the basis of the attributes available to and used by an observer. According to recent formulations, people differ in the number of cues that they accumulate prior to making a classifying (or naming) response. The purpose of this experiment was to test for relationships between amount of cue accumulation prior to recognition response and variables of personality as measured by the MMPI. Thirty-three male college student *S*s were shown stimulus object-name pairs to criterion and were then shown sequences of test figures providing more and more cue information relative to one of the previously learned objects. Each *S* was instructed, on each of four trials, to name the object when he recognized it. The longer he waited before responding, the higher his total recognition score. The only significant predicted relationship found was the correlation of .49 between *Pa* score and recognition score. It seems that the increasing distrustfulness and suspiciousness found in persons with higher *Pa* scores brings with it a tendency to give recognition responses only after a relatively large number of cues are accumulated. Significant positive correlations were also found between *Pd* and recognition score (.40) and *Sc* and recognition score (.38), but these results should be cross-validated.

REFERENCES

1. Binder, A. A statistical model for the process of visual recognition. *Psychol. Rev.*, 1955, **62,** 119–129.
2. Brackbill, G., & Little, K. B. MMPI correlates of the Taylor Scale of Manifest Anxiety. *J. consult. Psychol.*, 1954, **18,** 433–436.
3. Bruner, J. S. On perceptual readiness. *Psychol. Rev.*, 1957, **64,** 123–152.
4. Deese, J., Lazarus, R. S., & Keenan, J. Anxiety, anxiety reduction and stress in learning. *J. exp. Psychol.*, 1953, **46,** 55–60.
5. Eriksen, C. W. Some personality correlates of stimulus generalization under stress. *J. abnorm. soc. Psychol.*, 1954, **49,** 561–565.
6. Eriksen, C. W., & Davids, A. The meaning and clinical validity of the Taylor Anxiety Scale and the hysteria-psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50,** 135–137.
7. Grice, G. R. Discrimination reaction time as a function of anxiety and intelligence. *J. abnorm. soc. Psychol.*, 1955, **50,** 71–74.
8. Hilgard, E. R. The role of learning in perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: An approach to personality.* New York: Ronald, 1951.
9. Martin, B., & McGowan, B. Some evidence on the validity of the Sarason Test Anxiety Scale. *J. consult. Psychol.*, 1955, **19,** 468.
10. Miller, J. G. Unconscious processes and perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: An approach to personality.* New York: Ronald, 1951.
11. Moffitt, J. W., & Stagner, R. Perceptual rigidity and closure as functions of anxiety. *J. abnorm. soc. Psychol.*, 1956, **52,** 354–357.
12. Rockett, F. C. Speed of form recognition as a function of stimulus factors and test anxiety. *J. abnorm. soc. Psychol.*, 1956, **53,** 197–202.
13. Sarason, S. B., & Mandler, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47,** 810–817.
14. Sarason, S. B., Mandler, G., & Craighill, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47,** 561–565.
15. Smock, C. D. The influence of psychological stress on the "intolerance of ambiguity." *J. abnorm. soc. Psychol.*, 1955, **50,** 177–182.
16. Taylor, Janet A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53,** 303–320.

Received July 9, 1957.

HYSTERICS AND DYSTHYMICS AS CRITERION GROUPS IN THE STUDY OF INTROVERSION–EXTRAVERSION[1]

JOHN J. SIGAL,[2, 3] KOLMAN H. STAR,[4, 5] AND CYRIL M. FRANKS[6]
Institute of Psychiatry, Maudsley Hospital, University of London

Eysenck (3, 4, 5) has contended that hysterics are *extraverted* and neurotic, whereas dysthymics (Eysenck's designation for neurotics suffering from anxiety, obsessions, compulsions, phobias, or reactive depressions) are *introverted* and neurotic, with introversion–extraversion and neuroticism assumed as completely independent variables. Recently (7, 8, 10), he has grouped psychopaths with hysterics as extraverted neurotics. Although it is not a necessary consequence of the hypotheses in these papers that the hysteric–psychopath group is generally extraverted and the dysthymic group generally introverted, this assertion is integral to Eysenck's thinking.

If the postulates that hysterics and psychopaths are extraverted and dysthymics are introverted can be shown to have some basis in fact, the problem of obtaining satisfactory criterion groups with which to validate introversion–extraversion questionnaires (2) or objective tests that purport to measure this trait would be greatly simplified. One would only have to find relatively pure cases of hysteria or psychopathy on the one hand and dysthymia on the other to consider as valid, even in normal populations, any test that discriminated between the two.

On the basis of experimental and other findings, Eysenck (5) accepted Guilford's R scale and C scale as good measures of extraversion and neuroticism, respectively, in his sense of these terms. Both of these scales, however, have such undesirable elements as repetition of items with only slightly altered wording, items on which there are sex differences, and items that do not correlate with total scores (9). In order to overcome these difficulties and thereby to obtain purer measures to test deductions made from his theory, Eysenck constructed the Maudsley Personality Inventory (MPI), comprising an E scale (extraversion scale) and an N scale (neuroticism scale). Criterion groups for the item analysis consisted of extreme scorers on Guilford's R and C scales. Further details of the construction of the scales may be found elsewhere (9). It is sufficient to indicate here that, on the basis of a factor analysis of the items and experimental evidence, Eysenck considered the aim to have been accomplished.

The present study was undertaken as a direct validation of the E and N scales of the MPI within the Eysenckian framework. It was argued that if the E scale is a good measure of extraversion, then it should also discriminate between the hysteric–psychopath[7] group and a dysthymic group, the prediction being that the former would have a higher mean E score with the mean E score of normals falling between those of the two pathological groups. Furthermore, since the N scale is supposedly a measure of neuroticism, it was predicted that the two neurotic groups have a significantly higher mean score on this scale than normals but that they do not differ from each other.

METHOD

Sample

All the neurotic *S*s were adult inpatients of two mental hospitals. The majority were voluntary admissions. The chief psychologist of one hospital and a group of

---

[1] The authors are indebted to L. Minski and M. Desai, Superintendent and Chief Psychologist, respectively, of Belmont Hospital, for their permission to test patients, and to the psychiatrists of that hospital who facilitated the selection of suitable patients and cooperated in the testing program. G. A. Foulds, Chief Psychologist, Runwell Hospital, provided further invaluable assistance in the selection and testing of suitable patients.

[2] This research was carried out while J. J. Sigal was the recipient of a Post-doctoral Overseas Fellowship from the National Research Council of Canada.

[3] Now at the Hampstead Child-Therapy Clinic, London, England.

[4] This research was carried out while K. H. Star was the recipient of a grant from the Bethlem Royal and the Maudsley Hospitals' Research Fund.

[5] Now at Newcastle University College, New South Wales University of Technology, Tighe's Hill, NSW, Australia.

[6] Now at New Jersey Neuro-Psychiatric Institute, Princeton, N. J.

[7] Hysteric–psychopath is used in this paper to refer to a group consisting of *S*s with symptoms of hysteria or psychopathy or both, but with little if any dysthymic symptomatology.

psychiatrists at the other were asked to select those patients who had a predominance of dysthymic symptoms and a minimum of hysterical or psychopathic symptoms for one group, and the reverse for a second group. They were requested *not* to include any patients who had received courses of either ECT or coma or somnolent insulin within a period of eight weeks prior to testing, or patients who were receiving drugs in doses judged to have more than nighttime sedation effects, or brain-damaged patients. They were also asked not to include mental defectives or psychotics, since either might not be able to understand the instructions or to answer the questions properly.

Seventy-five patients were nominated as suitable, plus 64 charges of psychiatrists who offered all their patients as candidates but declined to select them further.

The case notes of each of these candidates were read by the authors, who arrived, independently, at a decision as to whether or not the candidate should be included in the sample and, if included, into which of the diagnostic groups he should be placed. In addition to rejection on the preceding criteria (Raven's Matrices or Mill Hill Vocabulary scores were available for all *S*s), candidates were rejected if their mother tongue were not English, if they had thinking disorders that might render it impossible for them to understand the instructions or the questions, or if they were not identifiable as either dysthymic, hysteric, psychopathic, or some combination of the last two. Diagnoses were based on symptomatological history only. After each of the three judges had arrived at an independent decision, a joint meeting was held in which these decisions were compared, and if unanimous agreement could not be reached to include a given *S* in a particular diagnostic group, the *S* was rejected.

Of the 139 candidates whose case notes were read, 81 were rejected. The remaining 58 were asked to volunteer or had already volunteered for the study, having first been told that they were at complete liberty to refuse to participate and that their results would in no way have a bearing on their treatment or hospital status. Five patients declined to participate, and one other discharged herself from the hospital before she could be tested.

The final sample consisted of 27 *S*s diagnosed as hysterics or psychopaths or having symptoms of both these disorders (the hysteric–psychopath group), and 25 dysthymics. The hysteric–psychopath sample was further divided (once more on the basis of unanimous agreement among the judges) into 15 *S*s clearly diagnosable as hysterics and 8 *S*s clearly diagnosable as psychopaths; the remaining 4 *S*s of the original group had mixed symptomatology.

The psychopaths as a group were typified by the following behavior: repeated lying, stealing, or forgery; repeated impulsive acts, and loss of temper. There was a tendency for these actions to recur despite many prison sentences or frequent dismissal by employers.

The primary symptoms of the hysterics were pains or paralyses in the hands, arms, or legs, pains in other parts of the body, or anorexia. In addition, there was a lack of manifest anxiety, depression, or dysthymic features.

The dysthymics typically had one or more of the following symptoms: feelings of depression, complaints of anxiety and tension, vague fears, difficulties in concentrating, giddiness, tremulousness, insomnia, obsessive thought, compulsive acts, or a variety of phobias.

In all cases the symptoms were without known, direct organic cause.

The group used as a normal control consisted of 213 male university students. Details of their selection are reported by Star (17). The fact that the control group consisted exclusively of males was not considered relevant because MPI items were selected in such a way as to eliminate sex differences, and Eysenck (9) found no significant sex differences in mean *E* and *N* scores in normal *S*s. That the control group was probably superior in intelligence to the neurotic groups was not relevant to the study either since Eysenck (5) has shown intelligence to be orthogonal to introversion–extraversion and neuroticism.

Procedure

The instructions and scoring method for the questionnaire were the same used by Star (17). One point was given for an extraverted or neurotic answer, and half a point when "?" was circled; the scores were, thus, exactly half of what they would be under Eysenck's scoring method. The range of possible scores was from 0 to 24 for both E and N scales.

RESULTS

The mean MPI *E* and *N* scores and standard deviations for dysthymics, hysterics, psychopaths, hysteric–psychopaths, and normal controls are shown in Table 1. Tests of differences between the various groups are given in Table 2. Because the *F* ratio for normals vs. dysthymics on the E scale was significant ($p < .05$), the Cochran-Cox test was used in testing the significance of the difference between the mean *E* scores of these two groups. All other *F* ratios were insignificant, so *t* tests were employed.

Extraversion. It can be seen from Tables 1 and 2 that the hysteric–psychopath group did not differ significantly from dysthymics on the E scale and, to three significant figures, had the same mean score as normals. Dysthymics, on the other hand, obtained a significantly lower mean score on this scale than normals. When hysterics and psychopaths were separately compared with dysthymics and normals, the only additional significant difference was between the mean *E* scores of dysthymics and psychopaths.

On the basis of these scores, there was no more reason for assuming that hysterics and psychopaths should be grouped together than there was for assuming that hysterics and dysthymics should be. Consequently, the

TABLE 1
MEAN MPI SCORES OF NEUROTIC GROUPS AND NORMALS

Diagnosis	n	*E* score		*N* score	
		Mean	*SD*	Mean	*SD*
Normal	213	12.63	4.44	11.62	5.65
Dysthymic	25	10.50	5.98	18.40	5.24
Hysteric–psychopath	27	12.61	4.98	14.41	6.38
Hysteric	15	12.07	4.86	12.03	6.41
Psychopath	8	14.88	3.42	18.81	3.12
Neurotic	52	—	—	16.33	6.09

TABLE 2
SIGNIFICANCE OF MPI SCORE DIFFERENCES

Groups	*E* score*	*N* score*
Dysthymic vs. hysteric–psychopath	NS[a]	$p < .02$
Normal vs. dysthymic	$p < .05$[a]	$p < .0005$[a]
Normal vs. hysteric–psychopath	NS	$p < .01$[a]
Normal vs. hysteric	NS	NS[a]
Normal vs. psychopath	NS[a]	$p < .0005$[a]
Dysthymic vs. hysteric	NS[a]	$p < .01$
Dysthymic vs. psychopath	$p < .05$[a]	NS
Hysteric vs. psychopath	NS	$p < .02$

* NS means $p > .05$.
[a] One-tailed test.

mean *E* score of hysterics and dysthymics combined (11.09 ± 5.50) was compared with that of the psychopaths. This difference failed to reach significance ($.10 > p > .05$) on a two-tailed test.

The best discrimination into hysteric–psychopath or dysthymic groups that could be obtained using the E-scale distributions resulted in a total misclassification of 37 per cent of the cases (33 per cent of the hysteric–psychopath group and 40 per cent of the dysthymic group).

Neuroticism. Normals had the lowest and dysthymics the highest mean *N* scores of the three principal groups; the mean of the hysteric–psychopath group fell between the two. Differences between normals and hysteric–psychopaths and between the latter and dysthymics were significant at beyond the .02 level. Hysterics alone did not score significantly higher than normals on the N scale but but did score significantly lower than dysthymics. Psychopaths alone obtained the highest *N* score of all groups, but their mean score did not differ significantly from that of dysthymics; it was, of course, significantly different from those of hysterics and of normals.

Although the difference between mean normal and neurotic *N* scores was statistically significant ($p < .001$), a cutoff at the score below which 90 per cent of the normal population fell still misclassified 21 per cent (8 hysterics and 3 dysthymics) of the neurotics. When the *N* scores were used as a means of distinguishing between dysthymics and the hysteric–psychopath group, and dysthymics were assumed to be the more neurotic, the total misclassification was 29 per cent. When the psychopaths were included with the dysthymic group (there is some justification for this since they did not differ significantly on the N scale) and the same cutoff score was used, the misclassification ($N = 48$) fell to 19 per cent (12 per cent of the dysthymic–psychopath group and 33 per cent of the hysteric group).

In order to explore the possibility that the results on the E scale were in some way determined by the differences between groups on the N scale, *S*s of the hysteric–psychopath group were matched on the basis of *N* scores with a group of dysthymics. In matching *S*s, a difference of one-half point between two members of a pair was allowed, and tables of random numbers were used for selection when two or more *S*s of one group might equally well have been matched with one *S* in the other group. Fourteen pairs were selected in this way. Mean *N* scores of 16.12 ± 6.23 and 16.25 ± 6.13 for the hysteric–psychopath and dysthymic groups, respectively, testify to the success of the matching procedure. The mean scores on the E scale of these matched groups were 10.96 ± 5.29 and 10.93 ± 5.37 for the hysteric–psychopath and the dysthymic groups, respectively. The difference is clearly not significant.

DISCUSSION

It is quite obvious that if the E scale is a measure of introversion–extraversion, and if the diagnoses of the *S*s are correct, then hysterics and psychopaths, considered as one nosological group, are not significantly more extraverted than dysthymics. Rather, they tend to be more like normals. Indeed, hysterics tend to have a lower mean *E* score than the normal group. There is a possibility that psychopaths are more extraverted than dysthymics, a finding supported by Hildebrand (13), but since the number of *S*s clearly diagnosable as psychopaths was small, a firm conclusion to this effect is not possible at the present time.

Because the procedure used in selecting the sample required unanimous diagnostic agreement among the judges, and because this judgment was based on symptomatology alone, it is unlikely that the failure to confirm Eysenck's hypothesis could be attributed to incorrect diagnosis. Certainly, these groups are diagnostically homogeneous when compared to groups used in previous studies (e.g., 11, 13, 15) in which different psychiatrists diagnosed the *S*s and in which experimenters expressed some reservations about the diagnoses. In addition, the descriptions of the groups according to a tally of the symptoms of the *S*s tested are consistent with the overt characteristics listed for the relevant disorders in the diagnostic manual of the American Psychiatric Association (1), and with Eysenck's earlier descriptions (e.g., 4, p. 122).[8]

It may be argued that psychiatrists rely on a global picture, rather than on symptoms alone, in making their diagnoses. However, in four independent studies (13, 15, 16, 18) in which the R scale served as a measure of extraversion, global psychiatric diagnosis was used, and in three of the four,[9] hysterics obtained essentially the same mean score as Eysenck's (9) normals, whereas in the fourth (15), they were significantly more introverted. It is unlikely, therefore, that the use of a global personality picture would have resulted in hysterics obtaining higher extraversion scores than normals in this study.

Nor is this the first time a measure of extraversion has failed to distinguish between hysterics and dysthymics. In Nichols' (16) data, the difference between hysterics and dysthymics on the R scale was not significant when the effect of sex differences was removed, and Martin (15) found the mean *R* score of hysterics to be insignificantly lower than that of dysthymics. Hildebrand (13), however, found a significant difference between anxiety states and hysterics on the R scale. Since no attempt was made to assess the interjudge reliability of diagnosis in any of these studies, it is possible that this introduced an unknown degree of experimental error in all of them.

Eysenck's assumption that neuroticism and introversion are independent in a neurotic population could not be tested within the present experimental design. However, the almost identical E-scale scores of the hysteric–psychopath and dysthymic groups matched on the N scale make it unlikely that even nonindependence of the two variables could account for the failure to find the expected differences between hysterics and dysthymics. Thus, the results of the present study are not without considerable support and are nowhere directly contradicted.

Three possibilities remain open: Either Eysenck is incorrect in postulating that hysterics are extraverted; or the E scale and, as a result, the R scale on which it is based are not good measures of introversion–extraversion; or both statements are true. If the first alternative were accepted, and if the N scale is considered to be an adequate measure of neuroticism, psychopaths and dysthymics might still be used as criterion groups in the measure of introversion–extraversion, since their mean *N* scores did not differ significantly but their mean *E* scores did. This must be considered as a tentative conclusion at best, both because of the small number of psychopaths in this study and because the psychopaths were all hospitalized and may, therefore, be representative of only one of the at least two experimentally definable groups of psychopaths (14). Acceptance of the second alternative would question the relevance of some studies (e.g., 12, 13) cited by Eysenck in support of his theories of personality (10) and politics (6) since, in these studies, the extraversion axis was rotated to the R scale, or the R scale had the highest loading on the extraversion factor. On the other hand, acceptance of either of the first two alternatives would not necessarily invalidate studies in which neurotic groups were selected on the basis of high and low *E* or *R* scores, provided the results are not interpreted to mean that the diagnostic groups from which *S*s were selected have been made diagnostically more homogeneous by this procedure. Using extreme scores on the E or R scales to screen further a group of hysterics or dysthymics will not necessarily result in the selection of more typical or purer groups of hysterics or dysthymics, but may result in selecting hysterics and dysthymics who perceive themselves as having extraverted or introverted patterns of behavior.

---

[8] We wish to thank H. J. Eysenck for making the manuscript of his most recent book available to us. Unfortunately, this study was carried out before the book had been completed and, as a result, it was not possible to attempt to objectify the descriptions given there for use as selection criteria. In any case, it would have been difficult, if not impossible, to obtain satisfactory interjudge reliability for the dynamic and structural personality constructs used by authors cited by Eysenck in describing the personality of the hysteric (e.g., hypertrophic egos in need of inflation, low centers overcharged at high tension, *die Heiligsten Gefühle*, etc.). It may be, therefore, that neurotics diagnosed as hysterics primarily on the basis of conversion symptoms, and lack of anxiety or other dysthymic features, are not what Eysenck would now call hysterics, or, at least, not the type of hysteric whose dynamics he wishes to explain in his more recent writings.

[9] One of these mean *R* scores, 37.29 ± 13.78, was obtained from an hysteric–psychopath group ($n = 14$) tested as part of another study (18), but the score is not reported there. H. P. Venables was most generous in making his original data available for these computations.

The results obtained suggest that the N scale discriminates between neurotics and normals with some success, although it did not differentiate significantly between hysterics and normals. To assess its value as a screening device in a practical situation, a more heterogeneous sample of neurotics than was used in this study would have to be tested. The results are, nonetheless, of some interest, for it would appear that the N scale is a better criterion for distinguishing between hysterics and dysthymics than is the E scale. Again, these findings are not without some support (11, 13). However, the evidence is not nearly as consistent in this respect as it was for the findings with the E and R scales, so these conclusions must be taken as tentative.

That hysterics did not differ significantly from normals on the N scale would seem to be irreconcilable with Eysenck's position unless the N scale is not a measure of neuroticism, since he argues that hysterics are both extraverted and neurotic. The consequences of this finding would parallel those mentioned in discussing the E scale.

SUMMARY

It follows from Eysenck's theory of personality that hysterics and psychopaths on the one hand and dysthymics on the other can be considered as typical groups of extraverted neurotics and introverted neurotics, respectively. To examine the usefulness of this concept for future validation of tests of introversion–extraversion, an apparently acceptable test of extraversion, the E scale of the MPI, and an apparently acceptable test of neuroticism, the N scale of the MPI, were administered to an hysteric–psychopath group ($n = 27$) and a dysthymic group ($n = 25$), selected on the basis of unanimous diagnostic agreement among three psychologists. The results suggest that either hysterics and dysthymics cannot be used in the described manner, or that the E and N scales do not measure introversion–extraversion and neuroticism, or that both statements are true.

The implications of these findings for Eysenck's theory and the experiments stemming from this theory were discussed.

REFERENCES

1. AMERICAN PSYCHIATRIC ASSOCIATION. Committee on nomenclature and statistics. Mental disorders. *Diagnostic and statistical manual.* Washington: Author, 1952.
2. CATTELL, R. B. Second order personality factors in the questionnaire realm. *J. consult. Psychol.*, 1956, **20**, 411–418.
3. EYSENCK, H. J. *Dimensions of personality.* London: Kegan Paul, 1947.
4. EYSENCK, H. J. *The scientific study of personality.* London: Routledge & Kegan Paul, 1952.
5. EYSENCK, H. J. *The structure of human personality.* London: Methuen, 1953.
6. EYSENCK, H. J. *The psychology of politics.* London: Routledge & Kegan Paul, 1954.
7. EYSENCK, H. J. A dynamic theory of anxiety and hysteria. *J. ment. Sci.*, 1955, **101**, 28–51.
8. EYSENCK, H. J. Cortical inhibition, figural aftereffect, and theory of personality. *J. abnorm. soc. Psychol.*, 1955, **51**, 94–106.
9. EYSENCK, H. J. The questionnaire measurement of neuroticism and extraversion. *Revista di Psicologia*, 1956, **50**, 113–140.
10. EYSENCK, H. J. *The dynamics of anxiety and hysteria.* London: Routledge & Kegan Paul, 1957.
11. FRANKS, C. M. An experimental study of conditioning as related to mental abnormality. Unpublished doctoral dissertation, Univer. of London, 1954.
12. GEORGE, E. I. An experimental study of the relation between personal values, social attitudes, and personality traits. Unpublished doctoral dissertation, Univer. of London, 1954.
13. HILDEBRAND, H. P. A factorial study of introversion-extraversion by means of objective tests. Unpublished doctoral dissertation, Univer. of London, 1953.

Bureau Ednl. ... Research
DAVID HA ... COLLEGE
Dated ....

14. Lykken, D. A study of anxiety in the sociopathic personality. *J. abnorm. soc. Psychol.*, in press.
15. Martin, Irene D. An experimental study of muscular tension in psychiatric patients. Unpublished doctoral dissertation, Univer. of London, 1955.
16. Nichols, E. G. The relation between certain personality variables and the figural after-effect. Unpublished doctoral dissertation, Univer. of London, 1956.
17. Star, K. H. An experimental study of "reactive inhibition" and its relations to certain personality traits. Unpublished doctoral dissertation, Univer. of London, 1957.
18. Venables, P. H. Changes in motor response with increase and decrease in task difficulty in normal industrial and psychiatric patient subjects. *Brit. J. Psychol.*, 1955, **46**, 101–110.

Received July 3, 1957.

FACIAL EXPRESSION AND SITUATIONAL CUES[1]

NICO H. FRIJDA

Amsterdam University

RECENT experimental research suggests that interpretations of facial expression of emotion can be made with reasonable correctness (6, 16, 20, 21). At the same time, knowledge of the situation that gives rise to the emotion expressed seems to modify interpretation (2, 20) and to add to its correctness (9, 14, 18, 20).

This paper tries to arrive at a clearer understanding of the relation between knowledge of the situation and recognition of expression proper. On the basis of some experimental evidence (3, 16, 22) and of theoretical considerations given elsewhere (6), it is supposed that the meaning of facial expressions is more general than the interpretations commonly attached to them. The greater specificity of interpretation is attributed to situational cues, which then can lead to greater correctness.

This view leads to the following hypothesis: If interpretations are analyzed, those aspects due to situational cues must be more specific than those due to expression alone. In operational terms: if situational cues are varied, the varying elements of interpretation are of a more specific nature; the constant elements of a more general nature. "General" is here defined as: a pattern of relational functioning (withdrawal, attention, tenseness, degree and nature of activity), regardless of the nature of the objects towards which this functioning is directed. "Specific" is defined as: any reference to an event in the inner or outer world, either explicit or implicit; or any reference to thought-content or to forthcoming behavior.

In experiments on the influence of situational cues (14, 18, 19, 20), interpretation of isolated facial expressions is compared with interpretation of pictures showing elements of the situation as well. This method does not seem the most fruitful road to theoretical analysis. When confronted with isolated expressions, the observer frequently imagines some situation around them which, of course, cannot be experimentally controlled (2, 6, 12, 15). A definite advance consists in placing the expressions in different contexts, as is done by Cline (3). A somewhat different approach is taken here, in which analysis can be pushed further.

METHOD

Procedure

In a classroom setting, two groups of *Ss* were presented with three or four photographs of an actress, projected successively on a screen (Fig. 1). *Ss* were given a response-sheet on which was printed a short description of the situation in which each photograph was supposed to have been taken. The first group of *Ss* received situation-description Series A, the second group Series B (see Table 1). Selection of pictures was guided by the desire for everyday, not exaggerated expressions.

The following instructions were read aloud:

> This is an experiment on recognition of facial expression. You will see some pictures, taken in the situations described on your response-sheet. Try to describe the person's feelings, emotional states, or whatever it may be. Do it in your own words, as freely as you choose and in as much detail as you wish.

A few seconds were allowed for reading the first description, the room was almost darkened, and Photo 1 projected on the screen. Unlimited time was permitted for all *Ss* to finish writing.

Subjects

The *Ss* were undergraduate students in psychology at Amsterdam University. There were 32 *Ss* in each group; 20 *Ss* only in each group saw also Photo 4. Slight fluctuations in numbers (see Table 2) are due to some illegible records.

Development of Categories for Interpretation

The interpretations given by *Ss* consisted mostly of rather extensive descriptions. For example, the following was given in response to Photo 1, Series A:

> She thinks of the time she was happily away with her lover; the picture she sees now brings back the recollection of it. That time is gone, and so is her lover. She is sadly remembering it all.

For purposes of analysis, a set of categories was devised by the author on the basis of the hypothesis to be tested and a perusal of the protocols. In Tables 2 and 3, summary descriptions are presented for all categories to which more than three interpretations were subsequently assigned in either Series A or B. "Presence of sad mood" (Table 3, Photo 1), for instance, stands for "the interpretation contains some shade of sadness or melancholy; that is, some element of a negative *mood*;

---

[1] This paper is based on part of a Ph.D. thesis made under the guidance of H. C. J. Duijker, to whom the author expresses his indebtedness.

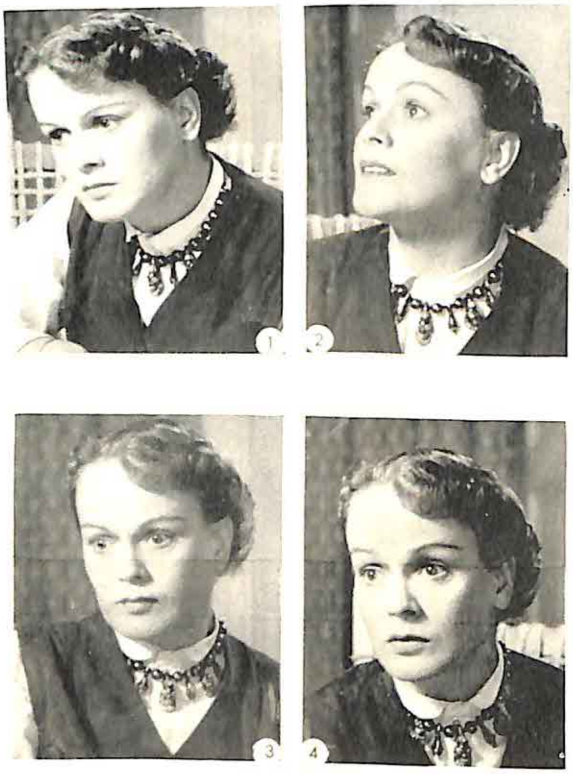

FIG. 1. FACIAL EXPRESSIONS USED IN EXPERIMENT

TABLE 1

ALTERNATIVE DESCRIPTIONS OF THE SITUATIONS PRESENTED TO *S*s

Series A	Series B
1. She is looking at a picture of a performance at the Folies Bèrgères (Paris Music Hall)	She has to solve a mathematical problem
2. Someone she likes very much says something sweet to her	She receives a serious reproach
3. She pulls the string of a dynamometer	She looks at an animal in the corner of the room
4. She receives a serious reproach	Someone she likes tells something sad about himself

In relation to the hypothesis to be tested, the categories were intended, on the one hand, to define those emotional elements which could be considered as directly influenced by situational components (as for instance sympathizing with the girls supposedly depicted on the Folies-Bèrgères picture). On the other hand, they sought to define the affective attitudes in more general terms, as for instance (again Photo 1): "Mention is made of some more personal aspect of e.g., 'she thinks of how good everything was some time ago and how different now'—but *not* the more active 'she is desperate.' "[2] It must be added that many of these categories—especially the most specific like "she is empathizing" (Table 2, Photo 1)—were established after their presence had been noted in the records.

TABLE 2

CATEGORIES DIFFERING IN NUMBER OF RESPONSES FROM EACH SERIES ASCRIBED TO THEM

Categories	A		B		p*
	(N = 32)		(N = 32)		
Photo 1:					
she is sympathizing or empathizing	10	31%	1	3%	<.01
mention of effort or taking pains	14	42%	22	66%	n.s.
mention of concentrated thinking	7	21%	17	51%	.03
she is wondering about something	7	21%	3	9%	n.s.
reactive feelings upon some outer event, like fear, disapproval	16	48%	2	16%	.01
feelings concerning own self, like frustration, insufficiency	5	15%	15	45%	.03
	(N = 31)		(N = 31)		
Photo 2:					
feelings of deeper, more personal kind	17	56%	3	10%	.01
she is reserved or reticent	3	10%	11	36%	.05
mention made of acceptance	19	63%	8	26%	.02
presence of positive mood	21	69%	2	6%	.01
presence of negative mood	0	—	9	30%	.01
mention of disagreeable surprise	9	30%	4	13%	n.s.
mention of pleasant surprise	1	3%	10	33%	.01
rejecting something, reaction of nonacceptance	1	3%	15	50%	.01
positive response, reaction of approval	18	60%	4	13%	.01
	(N = 30)		(N = 30)		
Photo 3:					
anxiety, fear mentioned	11	35%	18	58%	n.s.
startle, surprise mentioned	2	6%	16	53%	.01
effort or tenseness mentioned	26	84%	3	10%	.01
disgust, aversion mentioned	1	3%	9	30%	.02
attitude of reluctance	3	10%	0	—	n.s.
mention of interest, fascination	0	—	6	20%	.05
	(N = 20)		(N = 20)		
Photo 4:					
mention of compassion, sympathy	0	—	17	85%	.01
feelings of guilt, insufficiency, frustration mentioned	12	60%	5	25%	n.s.
rejecting something, reaction of nonacceptance	5	25%	2	10%	n.s.
fear mentioned	4	20%	3	15%	n.s.

* Two tailed test.

---

[2] Detailed descriptions of the categories (in the original Dutch and in English) have been deposited with the American Documentation Institute. Order Document No. 5557, from ADI Auxiliary Publication Service, Library of Congress, Washington 25, D. C., remitting $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 3

CATEGORIES WHICH SHOWED CORRESPONDENCE BETWEEN THE SERIES

Categories	A		B	
	$(N = 32)$		$(N = 32)$	
Photo 1:				
presence of sad mood	8	24%	10	30%
mention of effort or taking pains	14	42%	22	66%
feelings of deeper, more personal kind	17	51%	12	36%
mention of concentrated thinking	7	21%	17	51%
presence of painful shade in feelings	14	42%	10	30%
presence of activity of pondering	15	45%	12	36%
thinking as explicitly mentioned	22	69%	30	94%
	$(N = 31)$		$(N = 31)$	
Photo 2:				
presence of "open" attitude, attitude of intake or expectancy	20	66%	21	69%
mention made of acceptance	19	63%	8	26%
she is actively responding, or taking some stand	19	63%	19	63%
	$(N = 30)$		$(N = 30)$	
Photo 3:				
anxiety, fear mentioned	11	35%	18	58%
	$(N = 20)$		$(N = 20)$	
Photo 4:				
presence of sadness, negative mood	12	60%	12	60%
amazement, surprise mentioned	5	25%	4	20%
she is actively responding, or taking some stand	7	35%	8	40%
presence of deeper emotional involvement	14	70%	15	75%
feelings of guilt, frustration, insufficiency mentioned	12	60%	5	25%

feeling in which the 'deeper' levels of the person are involved." There is sometimes considerable overlap between categories due to the effort to catch subtler aspects as well as broader ones.

Application of Categories

The interpretations were assigned to the various categories by an independent judge who was ignorant of both the way these interpretations were obtained and the existence of two different situational series. He received the full descriptions of the categories and a complete list of all interpretations in their literal form (only omitting direct references to situational content, replacing, for instance, "the mathematical problem" by "a problem" or "Folies-Bèrgères" by "the spectacle"). Those from Series A and B stood in random order. The judge then read the responses to a given picture, considering one category at a time, recording the code-number of each response to which in his opinion this category applied. Since the relevant emotional aspects were often implicit in the descriptions, some degree of subjective evaluation was of course often necessary.

The author next separated the interpretations from Series A and B assigned to each category, and calculated their number and the percentage this made of the possible maximum (i.e., of N for the group). P values were computed by means of Fisher's Exact Treatment of 2 × 2 Tables (5, p. 96).

RESULTS

Table 2 clearly shows the influence of situational cues. In 17 categories, numbers of interpretations from Series A and B are significantly different. The situational cues sometimes lead to quite opposite impressions like "presence of negative mood" versus "presence of positive mood" (Photo 2). We thus can corroborate the findings of Cline (3), Munn (14), Sherman (18), Turhan (19), and the results on suggestion by Fernberger (4, 10) and Langfeld (12).

On the other hand, a number of categories showed no difference, or no significant one, while the emotional aspect concerned was present in at least 20 per cent of the interpretations in each series (Table 3). In as far as this correspondence between series may be considered an effect of the photographs, interpretation appears not to be determined by situational cues alone, as was suggested by Landis (11), Sherman (18), and Turhan (19).

Most of the categories of Table 3 seem to be more general, as defined above, than those of Table 2, as can be judged from Table 4. The more general categories show correspondence between series, while most of the less general ones show differences. In as far as this statement is valid, the hypothesis stated in the introduction is supported.

DISCUSSION

Table 4 shows, in the opinion of the investigator, that categories in Table 2 are not only more specific than those in Table 3 but also are specifications of the latter. For instance: "reactive feelings upon some outer event, like fear, disapproval" and "feelings concerning own self, like frustration, insufficiency" (Table 2, Photo 1) both can be "more personal feelings in which the deeper levels of the person are involved" (Table 3). Both are, or can be, forms of the latter—the first in relation to something external, the second to some evaluation of the self. Actually, 69 per cent of the interpretations assigned to "more personal feelings" also were assigned to the mentioned categories of Table 2 (the other 31 per cent being found in "sad mood" or being unspecified personal feelings). Examples which may illustrate this are: "She is afraid, feeling uneasy

TABLE 4
COMPARISON OF CATEGORIES

Showing Correspondence (From Table 3)	Situational Cue	Showing Difference (From Table 2)
Photo 1 More personal feelings in which the deeper levels of the person are involved; sad or painful feeling	A: referred to something external B: referred to evaluation of self	A: reactive feelings upon some outer event, like fear, disapproval B: feelings concerning own self, like frustration, insufficiency
Thinking, concentrated thinking and pondering (reflective activity)	A: concerning a human event B: concerning a problem posed	A: she is sympathizing or empathizing; she is wondering about something B: mention of concentrated thinking (with effort, taking pains)
Photo 2 Presence of "open" attitude, of intake or expectancy; (surprise)	A: towards something agreeable B: towards something disagreeable	A: acceptance; pleased surprise B: disagreeable surprise
She is actively responding, taking some stand		A: positive response, like approval; acceptance B: nonacceptance, rejecting something
(There is restraint in overt reaction)		A: presence of deeper, more personal feelings[a] B: she is reserved, reticient[a]
		A: positive mood B: negative mood
Photo 3 Anxiety and fear (involvement in some event)	A: in muscular activity B: in some apparition	A: effort and tenseness B: startle, surprise; disgust; interest, fascination
Photo 4 Presence of deeper emotional involvement; sadness	A: concerning evaluation of self B: concerning fate of other person	A: feelings of guilt, frustration, insufficiency B: mention of compassion, sympathy
She is actively responding, taking some stand	A: towards something disagreeable	A: rejecting something, reaction of nonacceptance

[a] These two categories were initially thought of as being more general and going to appear in Table 3.

that all this is performed in public," assigned to "more personal feelings" and to "reactive feelings"; and "Rising feeling of despair, fear to be unable to solve the problem," assigned to "more personal feelings" and to those "concerning own self."

A similar relationship can be found between the other categories, as indicated in Table 4. Some additions are made there, mentioned between brackets. The categories of Photo 3, Table 2, for instance, all can be considered "involvement." Actually, 47 per cent of the interpretations of Series B contain either a superlative or the word "anguish," indicating the involvement also present in effort and tenseness.

Some categories are present in both Tables 2 and 3. They seem to be fairly general but nevertheless "reinforced" by situational cues. It is not claimed here that the general categories are of the "correct" generality from a theoretical point of view.

We consider the categories of Table 3—and thus the correspondence between series—as due to the inherent meaning of the expressions. Evidence for this is in part indirect: there is frequently no evident relationship between these categories and the situational descriptions, nor seem these descriptions to contain a relevant common aspect. Neither sadness nor pondering (nor even sympathizing, Table 2, Photo 1) would have been expected when presenting a picture of presumably barely-clad girls ("she regrets not to be there, and in these thoughts there is something sad for her") or when presenting a mathematical problem ("because of what she is told to do she has been thinking of something else, something sad"). On Photo 2, about 60 per cent of *S*s see "surprise etc.," whether it be after a sweet remark or after a reproach. Many more examples could be added.

More direct evidence is obtained by comparing the categories of Table 3 with the expressive features themselves. It shows that they can be related unequivocally to each other.

Photo 1: Thinking, or more specifically pondering, is clear from the inactive pose and the not very intent or directed gaze. Sadness is clear from the downcast attitude. Some effort, be it from taking pains or from painful feelings, appears in the frown. Limitation of this activity to only the frown points at the "psychic" nature thereof: the presence of some deeper personal response.

Photo 2: The "open" attitude and the attitude of acceptance conform with the opened eyes. The presence of some active reaction appears in the hesitant way the mouth is held and also in something tense in the glance. The only incipient activity of mouth and head can be related to the "restraint" made responsible for the "reserve" and "deeper personal feelings" impressions (see Table 4).

Photo 3: Tenseness, fixity of glance, and backward movement are the principal components of fear and of a fairly intense degree of involvement at that.

Photo 4: The opened eyes again correspond to surprise. The evident activity of their opening and of the forward movement of the head constitute an active response. The contraction of the eyebrows makes it into something sad, something afraid even. Restriction of facial activity to this frown again suggests the presence of some deeper "inner" involvement.

Combining the foregoing, we may state as our general conclusion that facial expression represents only a general pattern: a general behavioral attitude, containing the state of attention and withdrawal of the person, the degree and state of his activity, and the degree to which he is involved in his own reaction; or the mere presence of some truly "inner" event. This general pattern closely parallels the functional value of the expressive pattern: of eyes, mouth etc., of approaching and withdrawing movement, and general or localized tension. The nature of these general attitudes seems to be related to Schlosberg's dimensions of emotion (17), which may well contain their more precise definition.

If this general attitude is taken to be the proper meaning of expression, each expression is unambiguous in these terms. However, the same general attitude can be present in very different emotions, because almost all emotional denominations include also situational reference and inner, not acted-out components. Therefore, interpretation in terms of emotions will show high variability. Moreover, emotions that contain the same general attitude will have identical expressions; therefore, they will be confused by the observer.[3] These inferences are completely in harmony with the grouping results of Woodworth (21, 22), the analysis of Schlosberg (16), and the experimental results of Cline (3), Vinacke (20), and the present investigator (6).

To specify the behavioral attitudes in terms of emotions, situational cues or situational suppositions based on experience are necessary. This is in harmony with Hebb's observations (8) and with his theoretical position, as well as that of Bruner and Tagiuri (1, p. 326).

SUMMARY

Four pictures of facial expressions were presented to two groups of observers, instructed to describe the emotional states present. Both groups also received short descriptions of the situations in which the photographs were supposed to have been taken, each group receiving a different set.

Interpretations of both series were compared. Differences were found, due to the situational cues, but similarities existed notwithstanding these cues. Analysis supported the initial hypothesis that expressive cues give rise to recognition of a general attitude or activity, the nature of which in terms of emotion can only be specified with the help of situational cues. This general attitude is taken as the intrinsic meaning of facial expression.

REFERENCES

1. BRUNER, J. S., & TAGIURI, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology*, Vol II. Cambridge, Mass.: Addison-Wesley, 1954.
2. CARMICHAEL, L., ROBERTS, S. O., & WESSEL, N. Y. A study of judgment of manual expression. *J. soc. Psychol.*, 1937, **8**, 115–142.
3. CLINE, M. G. The influence of social context on the perception of faces. *J. Pers.*, 1956, **25**, 142–158.
4. FERNBERGER, S. W. False suggestion and the Piderit model. *Amer. J. Psychol.*, 1928, **40**, 562–568.

---

[3] Frijda (6) showed to his *S*s a photograph of a fairly intense startle reaction following an explosion. Fifty-five per cent of *S*s interpreted: profound thought. Startle is reactive withdrawal upon an outer event; profound thought is active withdrawal because of inner events.

5. FISHER, R. A. *Statistical methods for research workers*. London: Oliver & Boyd, 1948.
6. FRIJDA, N. H. The understanding of facial expression in emotion. *Acta Psychol.*, 1953, **9,** 294–362.
7. FRIJDA, N. H. *Het begrijpen van gelaatsexpressies*. Amsterdam: G. A. van Oorschot, 1956.
8. HEBB, D. O. Emotion in man and animals: An analysis of the intuitive process of recognition. *Psychol. Rev.*, 1946, **53,** 88–106.
9. INUI, T. On the temporal structure in the judgment of facial expression. *Jap. J. Psychol.*, 1937, **12,** 465–482.
10. JARDEN, E., & FERNBERGER, S. W. The effect of suggestion on the judgment of facial expression in emotion. *Amer. J. Psychol.*, 1926, **37,** 565–570.
11. LANDIS, C. The interpretation of facial expression in emotion. *J. genet. Psychol.*, 1929, **2,** 59–72.
12. LANGFELD, H. S. The judgment of emotions from facial expressions. *J. abnorm. soc. Psychol.*, 1918, **13,** 172–184.
13. LANGFELD, H. S. Judgments of facial expression and suggestion. *Psychol. Rev.*, 1918, **25,** 488–494.
14. MUNN, N. Knowledge and judgment in the interpretation of facial expression. *Psychol. Bull.*, 1940, **35,** 324–328.
15. RUCKMICK, C. A. A preliminary study of the emotions. *Psychol. Monogr.*, 1921, Whole No. 136.
16. SCHLOSBERG, H. The description of facial expression in terms of two dimensions. *J. exp. Psychol.*, 1952, **44,** 229–237.
17. SCHLOSBERG, H. Three dimensions of emotion. *Psychol. Rev.*, 1954, **61,** 81–88.
18. SHERMAN, M. The differentiation of emotional expression in infants. I. *J. comp. Psychol.* 1927, **7,** 265–284.
19. TURHAN, M. [An experiment concerning the interpretation of facial expression.] *Istanbul Univ. Yayinlar*, 1941, No. 149. (*Psychol. Abstr.*, 1944, **18,** 1391).
20. VINACKE, W. E. The judgment of facial expression by three national-racial groups in Hawaii. I. *J. Pers.*, 1949, **17,** 407–429.
21. WOODWORTH, R. S. *Experimental psychology*. New York: Henry Holt, 1938.
22. WOODWORTH, R. S. & SCHLOSBERG, H. *Experimental psychology*. (Rev. ed.) New York: Henry Holt, 1955.

Received July 8, 1957.

THE ORGANIZING FUNCTION OF INTERACTION ROLES IN PERSON PERCEPTION[1]

EDWARD E. JONES
Duke University

AND R. DECHARMS[2]
University of North Carolina

Among the important factors conditioning our impressions of other persons is the nature of our interactions with them, or the purpose of the interactions. Most of our interactions with others are constrained by the relatively crystallized norms governing role relationships; they occur in a socially *structured* context. The fact that human behavior is thus regulated by sets of mutual expectations, and the fact that we interact as role participants as well as unique personalities, have important consequences for the stability of the social environment. It may be derived from these facts that we often need little or no information about the internal state or the personality characteristics of another person in order to carry on effective social intercourse with him (cf. Steiner, 4).

But roles are not only important in determining the quantity of information needed about another as a unique human being; they dictate the kinds of information that are needed to sustain the interaction effectively, and they determine the kinds of inferences to which this information gives rise. We assume that the first cognitive task at the outset of an interaction is that of mutual orientation to the roles of the participants. It is as if each participant must come to an initial decision (no matter how tentative or erroneous) regarding the nature of the social situation in which he is involved. Out of this decision evolves a set to attend to, and to employ in certain ways, the information provided by the other person. In a previous paper[3] (2), we have referred to this preparatory focussing as the adoption of an *inferential set*. These sets are assumed to have selective as well as transformative (organizing) functions in the process of forming impressions of personality.

There are undoubtedly a number of ways in which one could classify roles and the different inferential sets to which they give rise. We have found it helpful to segregate three general classes of interactive behavior: (*a*) interactions in which the dominant purpose is *facilitation of own goal attainment*, (*b*) interactions in which the perceiver is interested in making a *deterministic analysis of some other's personality*, and (*c*) interactions involving the explicit or implicit *application of social sanctions*.

Perhaps the largest proportion of interactions is covered by the rubric, facilitation of goal attainment. We often have more or less important goals which others may serve or obstruct in various ways. Others may be "gate keepers" in charge of some valued commodity or position, they may provide us with information when our greatest need is for structure and clarification, they may tell us that they like us when we are insecure about our social status, etc. All such situations of high involvement in some personal goal typically promote a type of inferential set that we have called *value maintenance*. The main criteria of such a set are: (*a*) a basic concern with making an approach-avoidance decision, i.e., "what is he doing to me that makes me want to approach or avoid him?" and (*b*) a readiness to attribute causality and responsibility to others for those of their behaviors which are related to our pursuits.

The deterministic analysis of an other's personality, secondly, gives rise to what we have called the *causal-genetic set*, since the

---

[1] This study was conducted in the program of the Organization Research Group of the Institute for Research in Social Science at the University of North Carolina. The O.R.G. is supported jointly by the Ford Foundation and the Office of Naval Research [Nonr-855(04)].

[2] Now at Washington University, St. Louis.

[3] The concept of inferential set was initially discussed in a paper presented by the senior author at the 1955 meetings of the APA. A more comprehensive theoretical statement of the consequences of role-induced sets was presented by E. E. Jones and J. W. Thibaut at the ONR Symposium on Interpersonal Perception at Harvard University, April 20–23, 1957.

interest is in exploring the historical roots and underlying conditions that have given rise to a set of behaviors. While the professional roles of psychiatry and clinical psychology provide the clearest representatives of this set, there are many interaction contexts—particularly those in which the individual is not personally involved—where the man-in-the-street is concerned more with the determinants than with the consequences of behavior. The main criteria of such a set are: (*a*) a tendency to deemphasize the stimulus person's responsibility for "causing" his own behavior in favor of a concern with the historical antecedents that help to explain *why* he acted as he did and (*b*) a tendency to assign priority to certain determinants and to make associations between perceived and inferred characteristics in line with an "implicit theory of personality" (1).

Finally, the application of social sanctions gives rise to the related cognitive activity of situation-matching. The *situation-matching set* involves evaluating the appropriateness or correctness of behavior in terms of the norms perceived as relevant to the situation. The mother who must decide whether to praise or punish, the supervisor who must pass on the merits of an employee, the jurist who must weigh the criminal's guilt—all of these are examples of roles in which this set is likely to predominate. Like the value-maintenance set, and unlike the causal-genetic set, the situation-matching set involves an evaluative decision. However, the distinguishing criterion of this set is that the evaluation is in terms of some applicable situational norm rather than in terms of personal goal facilitation and obstruction. A further characteristic is that the perceiver concerns himself more with the social consequences of behavior than with causal or antecedent conditions. Other things being equal, this attention to social consequences should lead to negative evaluation of the person to the extent that behavior is inappropriate, provided that the individual is perceived as capable of distinguishing right from wrong and as sufficiently well informed about the governing norms.

THE EXPERIMENT

An exploratory experiment dealing with the arousal of different inferential sets was planned in the foregoing theoretical context. The experiment involved the presentation of standardized information concerning a fictitious stimulus person, followed by ratings on a detailed set of personality variables. The stimulus person was always presented as an ex-prisoner of the Korean war who had signed a number of propaganda petitions that were highly critical of the United States. Two versions of the stimulus pattern gave different reasons for his signing the petition. In the first version, the stimulus person presented himself (via a tape-recorded interview) as an effective, well-adjusted person to whom the norms that were supposed to govern his behavior as a prisoner were very vague and equivocal. In the second version, the stimulus person was, on the contrary, an individual with a traumatic personal history and obviously low frustration tolerance. The norms were well defined in this second case and it was clear that the stimulus person was aware he had violated them.

Thus, the experimental *S*s heard a tape-recorded interview which either emphasized the vagueness of the norms or the inadequacy of the interviewee as reasons for his having signed the petitions. The *S*s were further differentiated in terms of their exposure to certain set-arousing instructions prior to the playing of the interview. By a procedure to be described in greater detail below, the *S*s were placed in various hypothetical role-relationships with the stimulus person—they were to imagine themselves as members of a medical-psychological board of inquiry (causal-genetic set), or members of a court martial screening board (situation-matching set), or merely as prospective buddies in a military situation (value-maintenance set). A final control group merely listened to the recording without prior instructions.

In planning the experiment, it had seemed to us that the *S*s in the causal-genetic set would tend to evaluate the second (clear norms—poor background) stimulus person more favorably than would the *S*s in the situation-matching set. The former *S*s should be more sensitive to information about the stimulus person's background and more likely to "excuse" him because of his unfortunate history. On the other hand, we felt that *S*s in the causal-genetic set would evaluate the first (vague norms—good background) stimulus person less favorably than the situation matching *S*s. The salient question for the latter *S*s concerns the extent to which the norms are clearly structured and clearly violated. No determinate predictions were made concerning the value-maintenance and the control *S*s, except that it was assumed that they would fall more or less in between the other two sets for each stimulus person.

METHOD

Subjects

Two hundred and sixty-two naval air cadets in their first week of preflight training served as *S*s. Each of the eight experimental groups was composed of a preflight class which met as a group for the experiment.[4]

Procedure

The experimenter (*E*) introduced himself as a civilian psychologist attempting to collect data from servicemen in all branches on their attitudes toward those who had committed crimes or misdemeanors while in prisoner-of-war camps. The *S*s were told that they would hear portions of an interview with a sergeant who had acknowledged signing propaganda petitions against the United States while a prisoner of war in Korea.

After this introduction, a brief portion of the bogus interview tape was played containing neutral information about the sergeant's service history. The *S*s then rated the sergeant on 32 traits, choosing one of seven scale points ranging from "extremely characteristic" to "extremely uncharacteristic" of the sergeant. These constituted the "before" ratings, since they were conducted before any of the independent variables were introduced.

After this initial rating, the *S*s were instructed to assume a specific role while listening to further portions of the interview and while forming a more definite impression of the sergeant as a person. The roles described by *E* constituted the attempted induction of the three inferential sets serving as the main independent variable of the study. The following instructions were given to the members of different groups.

1. Causal-genetic treatment. "Now, let's suppose for a while that each of you is a member of a medical psychological board of review set up to determine exactly why the prisoner did what he did—why Sergeant X signed the petitions. You're not on any kind of legal council or court martial jury, but you are specifically empowered to find out all the facts in the case in order to explain the prisoner's behavior in the prison camp."

2. Situation-matching treatment. "Now, let us suppose for a while that each of you is a member of a preliminary legal court of inquiry which has been set up to determine whether the prisoner should be tried for aiding the enemy and to decide what the specific crime should be with which the prisoner should be charged. It is up to you to decide whether there are grounds for a court martial. As in all cases of this type, you are to weigh the facts and act in the best interests of the service while at the same time being fair to the individual prisoner."

3. Value-maintenance treatment. "Now, let's suppose that you are a soldier like X and you are trying to decide for yourself whether he is a nice guy or not so nice a guy. In other words, let's suppose you'd like to find out whether he would make a good friend or not, whether this is a fellow you would personally like and enjoy being with."

Immediately following the attempted induction of sets, the *S*s were asked to examine nine standard questions which might be asked about the sergeant to obtain relevant information. They were instructed to rank the questions in terms of their importance and relevance in fulfilling the role prescribed for their group by the foregoing instructions. These questions were constructed so that each was intended to be of differential relevance to the three sets induced, and it was hoped that the *S*s' rankings would reveal the extent to which they had adopted the set that the instructions were intended to induce.

After ranking these questions, the *S*s were again reminded of their role, and the major portion of the interview tape was played. Half of the *S*s heard one pattern of information about the sergeant, and the remaining *S*s heard a radically different pattern of information.

After hearing the crucial portion of the taped interview, the *S*s made a final rating on the same 32 traits as in the "before" rating. Their concluding task was to answer a series of questions included to establish the validity of difference in stimulus quality between the two interview variations. These are the "after" ratings.

The Stimulus Patterns

We shall refer to the interviewee in the first stimulus pattern as "Sergeant Strong" and to the interviewee in the second stimulus pattern as "Sergeant Weak" (in all cases he was presented to the *S*s as a real person who would be called "Sergeant X" in order to protect his anonymity). Strong described himself in the interview as a forceful person of good background who was operating in terms of vague and equivocal norms. Weak described himself as a rather confused, neurotic individual who nevertheless made it plain that there were clear-cut norms proscribing any such activity as signing petitions. The following excerpts from the interviews are intended to point up these differences.

Sergeant Strong	*Sergeant Weak*
As far as I can remember, we were given no specific briefing on what we were and were not supposed to do and say when captured. There wasn't much official talk about being captured at all. . . .	Well, I guess everybody knew beforehand you were never to give anything but your name, rank, and serial number. We were warned against trying to outsmart the Reds by answering some of their questions and not others. They told us the easiest thing to do is not to say anything and not to sign anything . . .
Well, I've heard of that kind of thing (mental	Pretty soon, they began taking us to this old

---

[4] We are indebted to Wilse B. Webb and William O'Connor for their effective help in securing subjects.

torture), but they never pulled any of that on me. No, they talked to me and told me they didn't expect to get any military information from me, but that things would be easier if I cooperated in "ideological ways," whatever that means. I couldn't see much to lose, and I figured maybe I could make a better deal out of being a prisoner.	house for "interviews," they called them. Here they would keep us apart from each other for hours and keep telling us all kinds of false dope about each other until we didn't know what to believe. . . . By the time I signed the petitions I was pretty fed up. That stupid propaganda statement didn't seem very important to me at the time. I just wanted to see a face I knew.
(Strong didn't feel forced to sign the petitions) . . . Anyway, self-control is one of my strong points. Nobody's ever been able to make me do something I didn't want to do or didn't think was right at the time. I was vice-president of my high school senior class, but it wasn't too big a class—only 150 or so. I took the lead also in getting our school affiliated with a local athletic league . . . far as I know they still are.	Well looking back on it, you know, there really wasn't so much pressure, I just "folded up" for some reason, I don't know. I never was too much good at standing up to things. When I was trying to get through high school and support my family at the same time, the strain was just too much and I had sort of a nervous breakdown for a while. They almost made me a 4-F because of it.

In one sense, the task of the perceiver when confronted with either of these interview patterns is to resolve a conflict. In each case there is a reason implied for signing the petitions (i.e., vague norms or "cracking" under minimal pressure) in spite of the fact that it was clearly wrong to do so.

The Rating Scales

The 32 characteristics were chosen so as to represent eight relevant a priori clusters. Each cluster consisted of four traits, two of which were stated in the positive (favorable) direction and two in the negative. The sequence of traits was randomized and held constant during the two administrations. The *Candor* cluster consisted of the traits honest, frank, deceitful, and insincere. This cluster was included to check on the possibility that the sergeant might be perceived as dissembling in the interview. Generally, all *S*s saw both sergeants as candid and honest. The other clusters were *Patriotism* (loyal, dutiful, left-wing, subversive), *Resistance* (strong-minded, stubborn, easily swayed, eager to please), *Idealism* (high principled, dedicated, opportunistic, self-seeking), *Predictability* (dependable, predictable, inconsistent, unreliable), *Acceptability* (likeable, friendly, irritating, contemptible), *Thought Clarity* (clear thinking, penetrating, confused, mixed up), and *Adjustment* (mature, decisive, easily upset, maladjusted). While no rigorous attempt was made to check on the factorial purity of these clusters, the traits in each cluster were carefully examined and were observed to vary together in a coherent fashion.

Differences between cluster means as a function of stimulus person and/or set were uniformly tested for significance by an application of Snedecor's approximation technique (3) for performing analyses of variance with unequal cell entries. The smallest cell in the 4 by 2 design contained 19 *S*s. Because of a consistent correlation between "before" scores and change scores on each cluster, the means were adjusted for this regression and in all cases an analysis of covariance performed.

RESULTS

The Stimulus Persons

The first questions to be asked of the data concern the extent to which the investigators were successful in conveying to the *S*s the intended impression of the stimulus persons and the extent to which they were able to arouse different inferential sets. The stimulus persons in the present experiment were intended to differ both with regard to those background characteristics indicative of good versus poor adjustment and with regard to the clarity of the governing norms.

The postexperimental questionnaire provides one relevant source of data in "defining" the nature of the stimulus persons as perceived by *all* of the *S*s.[5] One of these questions was essentially a seven-point scale ranging from the statement, "X had a home background which gave him every advantage for coping with life," to the statement, "X had a home background which clearly left him unprepared to cope with life." Ratings of Sergeant Strong are significantly closer to the the first extreme than are ratings of Sergeant Weak. In fact, the data show very little overlap between groups.

A second question dealt with the perceived clarity of norms. The scale here ranged from the statment, "It was made perfectly clear to X in his training that he was not to aid the enemy in any way if captured," to the statement, "It was never made clear to X at all what he should do when captured." As expected, Sergeant Weak is rated significantly closer to the first statement than is Sergeant Strong. The degree of overlap is again minimal.

In response to other questions, *S*s rated Sergeant Weak as feeling more guilt than

---

[5] It should be noted that there were no significant differences in the answers to the postexperimental questionnaire which were a function of induced set.

Sergeant Strong and felt that Sergeant Weak deserved heavier punishment than Sergeant Strong. Finally, the *S*s tended to attribute Sergeant Weak's actions to "the kind of person he was" and Sergeant Strong's actions to "circumstances," relatively speaking. All of these differences were statistically significant ($p < .05$). Among other things, they indicate the importance of normative clarity in the attribution of guilt and decisions regarding punishment.

Further support for the assumption that the intended differences between stimulus persons were perceived in line with the information presented comes from the data on before-after changes in personality ratings. Change data concerned with generally perceived differences between the stimulus persons are presented in Table 1 (and in Fig. 1).

It is very clear that, regardless of set, the *S*s see Sergeant Strong as changing toward being more well adjusted and Sergeant Weak as changing toward being less well adjusted as a function of the crucial information in the script. The difference between the two sets of change scores on the *Adjustment* cluster is highly significant ($p < .001$). Not only is Sergeant Weak perceived as more maladjusted, but he is also seen as lower on the *Thought Clarity* cluster ($p < .001$). A related cluster of traits dealing with the capacity to resist influence provides additional support concerning the perceived differences between stimulus persons. There is a clear-cut decline in Sergeant Weak's perceived *Resistance* and a slight rise in Sergeant Strong's perceived resistance. The differences between the two are again highly significant regardless of induced set ($p < .001$). Finally, again in line with expectations, Sergeant Weak shows a clear-cut decline in perceived *Predictability*, whereas there is no discernible change as far as Sergeant Strong is concerned. Again, the difference between the two stimulus persons is highly significant ($p < .001$).

TABLE 1

BEFORE-AFTER CHANGE SCORES IN RATINGS

(Combined data for all *S*s regardless of set)

Trait Cluster	Stimulus Person	
	"Strong" ($n = 126$)	"Weak" ($n = 136$)
Adjustment	+3.00	−11.72
Thought Clarity	−1.22	−9.52
Resistance	+5.40	−5.84
Predictability	−.13	−6.88

Note.—A + sign means more of the characteristic was attributed after than before the critical information. All differences between Stimulus Persons are significant at $p < .001$.

In terms of these data, then, there is no question that the stimulus persons are seen as differing markedly on the expected dimensions of variation. Sergeant Strong is seen as relatively well-adjusted, strong-minded, and predictable. He is also correctly judged to be operating under vague and equivocal norms. Sergeant Weak is perceived to be a poorly adjusted person who is unpredictable and easily swayed but who is operating under rather clear-cut norms. It is of interest to note that these are the only dimensions along which the stimulus persons are seen to vary. That is, there are no differences observed in perceived *Acceptability*, *Idealism*, *Patriotism*, or *Candor*, which are the result of the stimulus patterns themselves.

The Sets

As reported in the procedure, a group of nine questions was presented to the *S*s after hearing the set-arousing instructions and before hearing the crucial stimulus information in order to provide a check on the success of our arousal attempts and, hopefully, to give the set more momentum and force in the experiment. The *S*s were instructed to rank-order the questions in terms of their relevance to the performance of their role. These questions had been preselected so that three were intended to reflect the concerns of *S*s in each of the three sets.[6] By averaging the assigned ranks for each set of questions it was possible to assign each *S* three scores expressing his stated interest in the three different kinds of set-relevant information.

The average rank scores for each trio of questions, for *S*s in each set condition, are presented in Table 2. The pattern of means indicates that we were at least moderately

---

[6] The following examples will serve to indicate the types of questions predicted to be differentially preferred by *S*s adopting the different sets: *causal-genetic*: "What kinds of personal problems did X have before ever entering the Army?"; *situation-matching*: "Did the Army say anything in training about signing the enemy's propaganda petitions?"; *value-maintenance*: "Did X think that what he was doing might help his fellow prisoners?"

TABLE 2

The Effectiveness of Set Induction in Terms of "Questions You Would Like to Have Answered"
(Mean Rank Data)

Set Induced	Question Trios								
	Causal-Genetic			Situation-Matching			Value-Maintenance		
	SP• Strong	SP Weak	Total	SP Strong	SP Weak	Total	SP Strong	SP Weak	Total
CG	12.10	12.87	*12.41* *SD* 5.11	18.21	17.32	17.85 *SD* 4.80	14.69	14.84	14.75 *SD* 3.87
SM	11.63	14.76	13.72 *SD* 5.33	15.32	14.02	*14.45* *SD* 5.32	18.05	16.15	16.78 *SD* 4.78
VM	12.27	13.47	12.92 *SD* 4.45	18.83	19.45	19.16 *SD* 4.14	13.90	12.12	*12.95* *SD* 4.17
All Sets			13.02			17.15			14.83
Comparison I	CG vs. SM			SM vs. CG			VM vs. CG		
P diff.	<.20			<.001			<.01		
Comparison II	CG vs. VM			SM vs. VM			VM vs. SM		
P diff.	<.50			<.001			<.001		

Note.—The lower the score, the more important the question trio.

• Stimulus person.

successful in inducing the various sets, as judged by the questions chosen. Those *S*s in the situation-matching treatment tend to rank the intended situation-matching questions lower than *S*s in other treatments ($p < .001$). Also in line with expectations, *S*s in the value-maintenance treatment tend to rank questions supposedly relevant to their set lower than other *S*s ($p < .01$). However, the tendency of the causal-genetic condition *S*s to prefer the questions relevant to their set is clearly not significant. It is difficult to know quite what to make of this finding, since everyone, regardless of set, showed a general tendency to give low ranks to the causal-genetic questions. It is not as though *S*s in the causal-genetic condition showed a preference for other-than-causal-genetic questions; on the contrary, they indicated a clear preference for causal-genetic questions. The lack of significance is due to the fact that *S*s in other set conditions also preferred the causal-genetic questions to a greater than expected extent.

Since there was some variability within conditions regarding the questions most likely to be ranked low, and because of the overlap between set conditions, a compromise was adopted in the analysis of the data. The data for each set were subdivided into two levels as a function of the *S*'s score in ranking the questions presumably appropriate to his set. Thus, those *S*s in the causal-genetic set who preferred the causal-genetic questions the most were assigned to Level A. Those who preferred the same questions to a lesser extent were assigned to Level B. In each set condition, the cutting point for levels was the median for rank scores on the set-relevant trio of questions. The effect of these levels was statistically examined in the analysis of each trait cluster, with the thought that only Level A data would be used if there proved to be significant effects of level. Actually, however, in the results to be presented below there were no such effects of level, and all *S*s are therefore considered in each analysis.

The Perceptual Consequences of Set-Arousal

Turning to the primary question underlying this research: what differences in the personality ratings, if any, may be attributed to the induction of different inferential sets? Our initial expectation was that *S*s in the situation-matching set would be more sensitive to information about the norms, whereas *S*s in the causal-genetic set would be more sensitive to information about the individual's personality and past history. The consequences of this "sensitivity" for ratings of the stimulus person are not entirely clear. There is no indication from the data that the causal-genetic *S*s differ from other *S*s in discriminating differences in the stimulus person's level of adjustment; i.e., they do not see Sergeant Strong as better adjusted and Sergeant Weak as more poorly adjusted than the other *S*s do. There is also no indication that *S*s in the situation-matching treatment are any more responsive to changes in information about the norms than are the other *S*s. In terms of the postexperimental questionnaire data, all *S*s regardless of set see the norms as quite clear-cut for Sergeant Weak and quite vague for Sergeant Strong. One problem in testing "sensitivity hypotheses" with data such as these is that the stimulus information itself is so potent that there is little room for the influence of sets on ratings which are based primarily on cues clearly provided by the stimulus person.

Turning from the selectivity to the transformative function of the inferential sets, what differences are there in the inferences drawn about the stimulus persons as a function of set? The results are presented graphically in Fig. 1. All *S*s, regardless of set, rate a given stimulus person similarly on the *Adjustment* and *Resistance* clusters (and also on the *Candor*, *Thought*, *Clarity*, and *Predictability* clusters which are not discussed here since they provide little additional information). Differences begin to emerge, however, when we look at the *Idealism*, *Patriotism*, and *Acceptability* clusters. *S*s in the situation-matching set rate Sergeant Strong as less acceptable and Sergeant Weak as more acceptable than the other *S*s do. This over-all effect, comparing the situation-matching *S*s with all other *S*s, is highly significant ($p <$ .005). It should be noted that this represents a reversal of the expected direction.

More or less in line with the findings on acceptability, the situation-matching *S*s also see Sergeant Strong as less patriotic ($p < .05$) and less idealistic ($p < .005$) than do the other *S*s. There tend to be slight differences in the other direction with regard to Sergeant Weak. That is, the situation-matching *S*s see Sergeant Weak as more patriotic than the causal-genetic and the control *S*s do, and as more idealistic than all other *S*s do. None of these latter differences is significant, but their direction makes the contrasting reaction of the situation-matching *S*s to Sergeant Strong all the more vivid.

No other differences in cluster means are attributable to set. A glance at Fig. 1 suggests, however, that the situation-matching *S*s differ from the *S*s in all other sets in the extent to which they make covarying ratings on adjustment and acceptability. That is, *S*s in the causal-genetic, value-maintenance, and control conditions would seem to associate *Acceptability* with *Adjustment* to a greater extent than the situation-matching *S*s, seeing the well-adjusted persons as more likeable than the poorly adjusted person. This same association does not seem to characterize the situation-matching *S*s. Table 3 shows the correlations between *Acceptability* and *Adjustment* ratings for each set, pooling across stimulus persons. While little significance should be attached to the absolute level of the correlations (because of this pooling procedure), the unique status of the situation-matching *S*s is borne out as expected. The correlation for the situation-matching *S*s is significantly lower than the

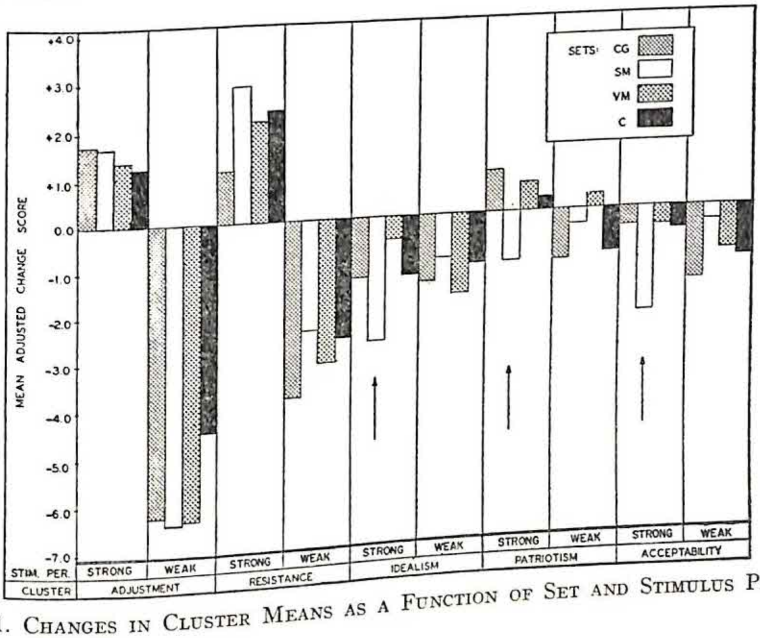

FIG. 1. CHANGES IN CLUSTER MEANS AS A FUNCTION OF SET AND STIMULUS PERSON

TABLE 3
CORRELATIONS BETWEEN SELECTED RATING CLUSTERS

Set	Correlation Between			
	Adjustment and Acceptability		Resistance and Acceptability	
	r	*P* diff.	*r*	*P* diff.
Causal-Genetic	+.428		.268	
		<.01		<.05
Situation-Matching	−.020		−.097	
		<.05		n.s.
Value-Maintenance	+.300		.030	

correlations for both causal-genetic and value-maintenance *S*s.

Correlations between rated *Resistance* and *Acceptability* are also included in Table 3. The striking difference here is that the causal-genetic *S*s definitely like the stimulus person more if they see him as having the capacity to resist influence and stand on his own two feet; the situation-matching *S*s, on the other hand, show no such association. The value-maintenance *S*s fall in between the others, somewhat closer to the situation-matching *S*s.

DISCUSSION

The experimental results make it quite clear that interaction roles (in this case, hypothetical) do indeed condition the adoption of different inferential sets. These sets, in turn, have important differential consequences for the utilization of information about the person being observed and evaluated. Our initial expectations about the specific nature of these consequences were not confirmed by the results, a fact which we believe emphasizes the importance of empirical research in this area. The results do fall into a meaningful pattern, however, and some attempt will here be made to interpret its significance. Each of these interpretive comments is offered only as a hypothesis that deserves extensive additional investigation.

The most dramatic set of findings concerns the extent to which the evaluative conclusions drawn by the situation-matching *S*s differ from the conclusions of the three other groups. All of our *S*s seem to have initially selected and registered the same relevant information about the stimulus person, but the subsequent inferences based on this information show clear and consistent differences across sets. The causal-genetic, value-maintenance, and control *S*s all seem to be victims of a rather undifferentiated halo effect. In their ratings, they characterize Sergeant Strong as well-adjusted, strong-willed, dependable, and likeable. Sergeant Weak, on the other hand, is seen as maladjusted, unreliable, easily swayed, and rather contemptible.

The inferences made by the situation-matching people are indeed strikingly different. They, too, see Sergeant Strong as well-adjusted and strong-willed, but they also see him as opportunistic, unpatriotic, and personally unacceptable (all relatively speaking). Sergeant Weak, on the other hand, is seen as poorly adjusted and easily swayed, but the situation-matching people like him and evaluate him more positively than the *S*s in any of the remaining groups. On the basis of these results, it might appear that the same characteristics which lead the first three groups to like or dislike the stimulus person lead the situation-matching *S*s to roughly the opposite conclusions.

In an effort to make sense of these data, and also in an attempt to re-establish contact with the theoretical propositions that prompted this experiment, we offer the following *ad hoc* interpretation. We had expected the situation-matching people to concentrate on the clarity of the norms and to evaluate the stimulus person positively when the norms are vague and negatively when they are clear. If we rely solely on the information provided in the stimulus scripts, almost precisely the opposite of this expected result occurred. But perhaps it is important to consider the possibility that our *S*s had rather clear ideas themselves regarding the relevant norms and that these played an important role in their judgments. That is, while we know that the *S*s perceived the intended difference between the stimulus persons with regard to the vagueness of norms, the cognitive awareness concerning norm clarity could not compete with their own views concerning the norms that would govern a prisoner of war. The experiment points up, in other words, the difficulty of experimentally specifying which norms are relevant in a situation where certain structured norms already exist.

It is useful, and it makes sense in terms of theoretical predictions, to assume that the situation-matching people are more preoccupied with the offense itself than any of the other subject groups. That is, the offense of signing propaganda petitions is always there, serving as a backdrop against which additional information will be evaluated. As a consequence, the offense is linked more directly to the attributes of the stimulus person. If he is seen as well-adjusted, strong-willed, etc.,—well, then, he should have known better than to have signed such petitions. How could one like or approve of such a person who deliberately violates the norms? The maladjusted stimulus person, however, is clearly less responsible for his transgressions and thus there is less of a tendency to dislike him. The crucial intermediate link in their line of reasoning seems to be the concept of *responsibility* (which is not to be equated with "intent"). Responsibility for a counter-norm offense leads to rejection by the situation-matching perceiver. The same offense committed unwittingly or under duress does not lead to rejection of the person.

In contrast to this preoccupation with the consequences of the stimulus person's behavior, *S*s in the causal-genetic and value-maintenance conditions seem to be responding just as if no offense had been committed. The stimulus person who [conveys] the image of a virile, well-adjusted male from a good middle-class background is seen as more likeable and dependable than the neurotic, uncertain, and inadequate person in our second interview. The assumption that the causal-genetic and value-maintenance *S*s like the stimulus person because he is well-adjusted and strong-willed is supported by the significant correlations between adjustment and acceptability and resistance and acceptability clusters. The situation-matching people, on the other hand, show insignificant *negative* correlations between these same sets of clusters.

The judgmental reactions of the causal-genetic *S*s merit some discussion since they, too, were out of line with our initial expectations. We had supposed that because of their sensitivity to the stimulus person's background and personality weaknesses, they would tend to make allowances for the neurotic, ineffective prisoner but not for the other stimulus person. Our data show, however, that the causal-genetic judges evaluate the stimulus person positively or negatively as a direct function of perceived level of adjustment. Perhaps it would be stretching things a bit to conclude that "a little psychology is a dangerous thing," but it does appear that the *S*s in the causal-genetic set are reaching the kind of evaluative conclusions that are supposedly inhibited in the process of training for diagnostic and psychotherapeutic work. It appears that a concern with the etiology and dynamics of behavior does not automatically guarantee a sympathetic response to the neurotic and the immature. In fact, our results suggest that without the concurrent inculcation of specific attitudes and values which might give rise to such sympathy, the reverse may indeed occur. The association between being perceived as well-adjusted and likeable is higher in the causal-genetic group of *S*s than in any other.

In conclusion, the present investigation shows that there are significant differential effects associated with the arousal of various inferential sets. These effects are revealed when standard information is transformed in different ways toward different conclusions in the inference process. The present study also shows how difficult it is to predict beforehand the facts that will be considered relevant and the precise course that inferences will take as a function of the set aroused. The pattern will not begin to take on clear structure until we have sampled the consequences of set-arousal in a variety of different situations. In this way the generalization value of the present results can be empirically established.

SUMMARY

The present experiment attempted to explore the effects of three induced "inferential sets" on the perception and evaluation of two radically different stimulus persons via a bogus tape recording. The *S*s were asked to assume one of three different roles in evaluating the actions of a prisoner of war who had apparently given aid to the enemy. The role descriptions were intended to induce one of the following sets: (*a*) causal-genetic, (*b*) situation-matching, or (*c*) value-maintenance. Each subject rated the stimulus person to whom he was exposed on the same set of 32 personality traits before and after hearing the crucially varying information about him.

The results show consistent and dramatic differences in the way the two stimulus persons

were perceived. The intentionally varied information was picked up equally well by all *S*s, regardless of set. The induced sets determined, however, how this information was interpreted. Compared to all other *S*s, those with a situation-matching set judged a strong-willed person operating under vague norms to be less personally acceptable, less patriotic, and more opportunistic. Another group of *S*s in the situation-matching set tended to rate a weak-willed person under clear norms more positively on these same characteristics than *S*s in the other sets. These results were interpreted as showing that situation-matching *S*s processed the same information as all others, but concentrated on the responsibility of the stimulus person for his actions, whereas *S*s in the other set conditions tended to accept or reject the stimulus person in accord more with the general cultural stereotype of a likeable person. The offense committed seemed to play a less decisive role in their inferences.

REFERENCES

1. CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.
2. JONES, E. E., & DECHARMS, R. Changes in social perception as a function of the personal relevance of behavior. *Sociometry*, 1957, **20**, 75–85.
3. SNEDECOR, G. W. *Statistical methods.* Ames: Iowa State, 1946.
4. STEINER, I. D. Interpersonal behavior as influenced by accuracy of social perception. *Psychol. Rev.*, 1955, **62**, 268–274.

Received July 10, 1957.

DEPRIVATION AND SATIATION OF SOCIAL REINFORCERS AS DRIVE CONDITIONS

JACOB L. GEWIRTZ[1]
National Institute of Mental Health

AND DONALD M. BAER[2]
University of Washington

In an earlier study (6) we received what appeared to be an affirmative answer to the question: Are there social drives that respond to reinforcer deprivation as do the primary appetitive drives? In this investigation we extend the question, asking in addition if the behaviors maintained by social reinforcers are responsive also to a condition of relative satiation for such reinforcers. Children are again employed as subjects (*S*s).

In the earlier study it was found that when an adult made words and phrases like "Good!" and "Hm-hmm" contingent upon an arbitrarily chosen response in nursery school children, that response was reinforced (i.e., conditioned). This effect was similar to that found in several other studies using verbal stimuli appealing to the concept of *approval* as reinforcers (e.g., 2, 9, 12, 16). It was found, in addition, that this reinforcing effect of approval could be increased when the children experienced a preceding 20-minute period of social isolation, relative to its effectiveness for the same children when they had not been isolated. While this result held primarily for boys tested by a female (rather than male) experimenter, other aspects of the data clearly supported the equating of isolation to the deprivation of social reinforcers: *social isolation increased reliably the reinforcing power of adult approval for children as a positive function of the degree to which they typically sought such approval in other settings.* Approval was taken to be representative of the reinforcers which control the purely social initiations made by children to adults.

Deprivation implies a period of unavailability of a given reinforcer, which results in an increase in behaviors for it; *satiation* implies a period of availability of a reinforcer, sufficient to effect a decrease in behaviors for it. Thus, deprivation and satiation represent two statements of a single concept, a dimension characterized by the relative supply of a reinforcer in the recent history of an organism which determines the incidence of behaviors for that reinforcer. The concept of deprivation-satiation has considerable precedent as a drive operation in general behavior theory (e.g., 11, 14), where it has been employed to order contemporary conditions which account for variance in reinforcer effectiveness. As such, drive is generally defined as the functional relation between deprivation (or satiation) for a reinforcer and responding for that reinforcer. Further, concepts like deprivation have been somewhat loosely applied in a number of speculative formulations of the antecedents of certain social behaviors (e.g., 1, 8, 13, 15). Hence, laws relating social deprivation as an empirically defined dimension to certain basic characteristics of social behaviors would have considerable integrative value (5). But first, the experimental operations of deprivation and its inverse, satiation, must be implemented effectively in social terms.

In the earlier study cited (6), a beginning attempt was made to implement social deprivation: brief social isolation of a child was equated to a condition of deprivation of all social reinforcers (including approval), and the differential effects of that condition and of a comparison nonisolation (nondeprivation) condition were reflected in the reinforcing effectiveness of an adult's approval. The present study represents an attempt to simplify (and replicate) the social deprivation operation of the earlier study, as well as a beginning in the direction of establishing an operation of social satiation. Satiation will be equated to a condition in which an abundance of approval and social contact is supplied to a child by an adult. Experimental operations implementing the conditions of deprivation and of satiation for a class of social reinforcers are both reflected against an intermediate or nondeprivation (nonsatiated) condition. The hypothesis is that these conditions should enhance the effectiveness of the reinforcer in the order: Deprivation > Nondeprivation > Satiation.

---

[1] This study was carried out when the senior author was on the faculty of the University of Chicago, and was facilitated by a grant given to him by the Social Science Research Committee of that institution. The writers acknowledge with gratitude the discriminating assistance of Chaya H. Roth.

[2] At the time of this study, the junior author was a Public Health Service Predoctoral Research Fellow of the National Institute of Mental Health at the University of Chicago.

METHOD

Sample

The *S*s were 102 middle-class children selected from the classes of the first and second grades of a university laboratory school and randomly assigned to experimental conditions. Sixteen were Negro, the remainder white. Their mean age at the time of testing was seven years, six months, with a range from six and one-half to nine years. Their mean and median Stanford-Binet IQ score was 127 (the scores of only 3 *S*s were below 100). One-half of the *S*s under each condition were girls and one-half boys. The *S*s were selected by their teachers according to the order in which their names appeared on alphabetical class lists. No *S* refused to participate.

Independent Variable

Deprivation. Seventeen boys and 17 girls were subjected to a condition of social isolation before playing the game. Each of these *S*s was introduced to the experimenter (*E*) in the classroom by the teacher. The *E* was a young woman in her early twenties.[3] She walked with *S* a distance of several hundred feet through the school corridors to the experimental room. During this walk, *E* responded to *S*'s questions and comments only when necessary, and maintained a somewhat distant but not unfriendly manner at all times. Upon reaching the experimental room, *E* showed *S* around the room, seated him and told him that someone else was using the game which he was to play and that she would have to fetch it but would be back in a little while. *E* then left the room and went (unobserved) to an adjacent observation booth from which *S* was observed during his isolation. She returned after 20 minutes with the toy. The game was then played in the usual fashion. The *S*s, who occasionally accompanied adults in the school setting for tests, and who had experience in awaiting their turns, all accepted this condition without question.

Nondeprivation. Seventeen boys and 17 girls were subjected to a condition of nonisolation, i.e., they played the game immediately upon their arrival at the experimental room. (Since this condition represents both relative nondeprivation and relative nonsatiation for approval, it served as an intermediate control condition for the other two conditions.) Until they entered the experimental room, *S*s in this group were treated identically as were *S*s in the Deprivation group.

Satiation. Seventeen boys and 17 girls were subjected to a condition of relative satiation for approval from the adult before playing the game. Again, each of these *S*s was introduced to *E* in the classroom by the teacher, but during the walk to the experimental room, *E* maintained a very pleasant and interested attitude toward *S*, responding to all details of his comments and questions, asking questions to draw out more of *S*'s conversation, and generally approving of anything about *S* which might reasonably be praised or admired. Upon reaching the experimental room, *E* showed *S* around the room, seated him, told him that the game was in use elsewhere and that she would go fetch it in a little while when it would be free. She suggested that meanwhile *S* might like to draw pictures or cut out designs, and proffered the essential materials. Then, for 20 minutes *S* drew or cut out designs, while *E* maintained a stream of friendly conversation with him, inducing him to talk about himself if he did not do so naturally. The *E* alternated her praise and admiration of whatever *S* did with whatever he said about himself, all in an appropriate fashion, and attempted to dispense 30 such reinforcers during the 20-minute satiation period at an approximate rate of three every two minutes. In fact, *E* dispensed to the boys an average of 31.6 such reinforcers ($\sigma = 6.9$) and to the girls an average of 28.2 reinforcers ($\sigma = 8.1$).

The Game Setting

Following the experimental treatment, the central task for *S* (which was the same for all three groups) was to place marbles into either of the two holes of the toy (shown in Fig. 1) while *E*, who sat beside him, looked on. (This procedure was identical to that employed in the earlier study (6) and is only summarized here.) The *E* observed *S*'s play for a "baseline" period of four minutes, during which *no* reinforcers were dispensed. Meanwhile, *E* responded to any of *S*'s comments and questions in a friendly but brief manner. Without pause, the baseline period was followed by a 10-minute test of reinforcer effectiveness. That is, *E* proceeded at this point to dispense the reinforcer, consisting of words like "Good," "Hm-hmm," and "Fine," according to a schedule incorporating several, successively higher, fixed ratios, whenever *S* dropped marbles into the *correct* hole, defined as that preferred *least* during the last (fourth) minute of the baseline period.

Dependent Variables

Reinforcer effectiveness score. The determination of the effectiveness of approval as a reinforcer was made from the "game" which followed immediately the treatment condition (i.e., the 20 minutes of Deprivation or of Satiation, or in the Nondeprivation condition, the arrival at the experimental room). The basic data were the numbers of marbles dropped in the correct and incorrect holes during each minute of play. These generated four pairs of frequency scores for the baseline period, and 10 such pairs of scores for the reinforcer effectiveness test period. For each minute of play, the *relative frequency* of a correct response was the ratio of correct responses to total responses, i.e., # correct/(# correct + # incorrect). The score employed as dependent variable was taken as the difference, for each

---

[3] This *E* served as one of the two *E*s in a companion study (6).

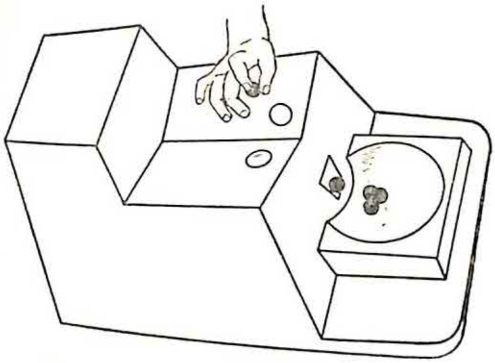

FIG. 1. A SCHEMATIC DIAGRAM OF THE EXPERIMENTAL APPARATUS (THE "GAME")

S, between the relative frequency of correct responses in the last (fourth) minute of the baseline period (when no reinforcers were dispensed) and the median relative frequency of correct responses of the 10 reinforcer test period minutes. This *reinforcer effectiveness score* represents the *gain* in relative frequency of the correct response attributable to the social reinforcer provided by the adult's approval.

Spontaneous social initiations. In addition, the verbal and purely social initiations of the child to the adult were tallied, but *only during the baseline period of the game*, before she began to dispense approval. The *E* treated the game as *S*'s central task and, while permissive, generally discouraged lengthy initiations on the part of *S*, suggesting, when necessary, that they could converse at length after the game. Where a reply to an initiation of *S* was required, *E*'s response was always friendly, yet brief, leaving responsibility for continuing that interaction sequence in *S*'s hands. All such responses by *S* were tallied by an observer in an observation booth. These responses took the form most frequently of Comments, less frequently of Questions, and least frequently of overt Attention-seeking.[4] They could be expected to represent behaviors by *S* for a variety of social reinforcers, in addition to approval. *Comments* were casual remarks which usually required no formal response from *E* (e.g., "We're going away for the holiday"). *Questions*, which were also casual, required only brief, token replies from *E* (e.g., "Do you think it will rain?"). When making a comment or asking a question, *S* typically continued responding in the game. *Attention-seeking* included responses characterized by urgency designed for active notice from *E*. Typically, *S* would pause to direct his complete attention to *E* while awaiting a response (e.g., "Watch me put the marble into this hole!"). Because of their generally infrequent occurrence in this study, these three behavior categories are scored in two different ways to produce two relatively independent response indices for the purpose of analysis:

a. The sum of the frequencies in the three social response categories, Comments, Questions, and Attention-seeking, which weights those categories in proportion to their frequency of occurrence (i.e., Comments contributes most and Attention-seeking least to that index).

b. The score of a *cumulative Guttman-like scale* (10), indexing what appears to be the intensity of social contact, formed similarly as in (3) when the frequency scores for each verbal response category were dichotomized for each *S* into gross response alternatives, zero or nonzero frequency of occurrence.[5] A scale score of 3 indicated that an *S* had exhibited some of all three behaviors; a score of 2 indicated that some Comments and Questions were exhibited, but no Attention-seeking; a score of 1 indicated that only Comments were exhibited; and a score of zero indicated that no social response of any type was exhibited. Hence, the scale scores are weighted in favor of the less frequently occurring behavior categories, Questions and Attention-seeking.

RESULTS

The hypothesis advanced is that the mean reinforcer effectiveness scores for the treatment groups would rank in the order Deprivation > Nondeprivation > Satiation. To test this, two relatively independent statistical procedures are followed: the first is sensitive to the rank-order of the treatment means but does not take account of the degree of overlap between the distributions upon which the means are based; and the second, which takes such information into account and allows parametric statements about the means, is somewhat less sensitive to the rank-order hypothesis.

Reinforcer Effectiveness Score

Rank-order of the means. In Table 1, it is seen that the rank-order predicted for the three means is found separately for Boys and for Girls. The sex variable may be taken as an independent replication of the experiment. The theoretical probability of obtaining a predicted rank-order of three independent means is $\frac{1}{6}$, and with one independent replication the probability of obtaining two such orders is $\frac{1}{6} \times \frac{1}{6}$ or $\frac{1}{36}$. On this basis, we can conclude that the null hypothesis, that all six rank-orders of the three means are equally likely, is rejected at $p < .03$. Hence, the alternative hypothesis, that the predicted rank-order of the three means prevails, is accepted.

---

[4] These three observation categories are defined in detail in (3). High agreement between observers was found there on these behavior categories, and they entered into similar patterns of relationship with the independent variables. Moreover, they were found in (4) to have high loadings on a single factor and appeared to involve an active attempt to gain or to maintain the adult's attention, perhaps differing in the degree to which their initiations for attention were overt or direct.

[5] Of all 102 *S*s in the four-minute baseline period, the proportion showing nonzero category frequencies was .71 for Comments, .53 for Questions, and .15 for Attention-seeking. When the response categories were arranged in the order of their decreasing popularities (the proportion of *S*s exhibiting nonzero frequencies), it was concluded that, for present purposes, a satisfactory three-item Guttman-like scale was produced: only .03 of the 306 responses produced scale errors, the few scale errors found appeared random, and every observation category contained far less error than nonerror. Of the 68 *S*s in the Deprivation and Nondeprivation groups, five each produced one scale error in his response pattern. Such response patterns were assigned scale scores corresponding to the higher pattern which would have been attained had there been no scale error.

Regression analysis. For the purposes of the regression analysis (and for theoretical reasons as well), the three treatment conditions are considered as points along a single continuous dimension representing the degree of social deprivation. The analysis then proceeds from three working assumptions: (*a*) the units separating the treatments along this deprivation dimension are roughly equal in size, (*b*) their relationship to the reinforcer effectiveness means is linear, and (*c*) it is the *pattern* of the three means that is relevant to the rank-order hypothesis advanced, not the contrasts between mean pairs representing adjacent points on the postulated treatment dimension. On this basis, it appears most efficient to carry out the analysis of variance according to a regression model: in essence, to analyze the regression of the reinforcer effectiveness score means on the treatment dimension.

TABLE 1

TABLE OF MEANS

Sex	Deprivation (D)	Satiation (S)	½ (D + S)	Non-deprivation
Boys	.36*	.18*	.27**	.21*
Girls	.33*	.07*	.20**	.23*
Combined	.34**	.13**	.23***	.22**

* 17 cases; S.E. of a mean difference is .068.

** 34 cases; S.E. of a mean difference is .048.

*** 68 cases.

TABLE 2

SUMMARY OF THE ANALYSIS OF VARIANCE OF THE REINFORCER EFFECTIVENESS SCORE

Source of Variation	*df*	Mean Square	*F*
Boys vs. Girls (Sex)	1	.0412	.524
Treatments			
Total	(2)	(.4032)	(5.123*)
Deprivation (D) vs. Satiation (S)[a]	1	.8031	10.200**
Nondeprivation (N) vs. ½ (D + S)[b]	1	.0034	.043
Sex × Treatment interaction			
Total	(2)	(.0356)	(.452)
Sex × D vs. S	1	.0268	.341
Sex × N vs. ½ (D + S)	1	.0444	.564
Error (within groups)	96	.0787	—

[a] Due to regression.

[b] Due to the deviation of the *N* mean from the regression line.

* $p < .01$.

** $p < .005$.

The reinforcer effectiveness scores of the *S*s were classified in the six cells of a 3 × 2 factorial, there being three treatments and two sexes of *S*s. After a Bartlett test indicated that the variances were homogeneous and an examination of the data suggested that the other assumptions underlying the analysis appeared to hold, the data were subjected to an analysis of variance. Table 2 indicates that there is no over-all sex difference, and that the treatment conditions do not affect the sex groups differentially. From Fig. 2, it is seen that the group means corresponding to the three levels of the experimental variable rank-order according to the hypothesis. It is seen in Table 2 that the Deprivation vs. Satiation comparison is reliable and in the order predicted. In the case of three means, this comparison represents the regression of those means on the treatment dimension. A Nondeprivation vs. ½ (Deprivation + Satiation) comparison, which represents the deviation of the Nondeprivation mean from the regression line, indicates that that mean falls remarkably close to the regression line. Hence, the regression predicted was found. (This result provides presumptive evidence only for the effectiveness of either of the extreme treatments, Deprivation or Satiation, relative to the intermediate Nondeprivation treatment.)

Approval as a Reinforcer. A question of interest is whether the reinforcer dispensed was effective as such under Nondeprivation, which would be comparable to the usual experimental case under which the effectiveness of the reinforcer would be examined, as well as following Satiation, when the reinforcer was relatively least effective in this experiment. The hypothesis is tested that the mean reinforcer effectiveness score under Nondeprivation is zero, employing an error term based only upon those 34 scores and a one-tail *t* test. That mean score (.22) is found to be reliably greater than zero

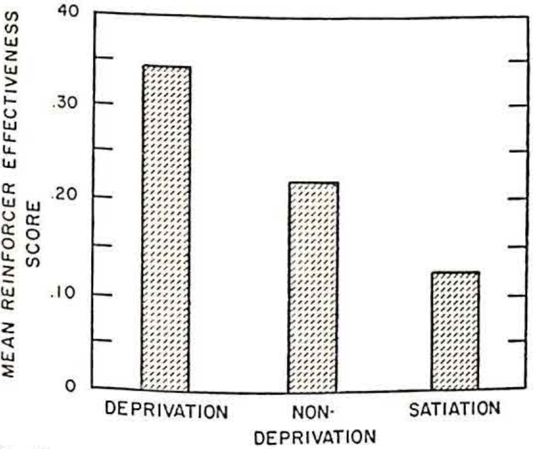

FIG. 2. A PICTORIAL REPRESENTATION OF THE MEANS FOR THE THREE LEVELS OF THE EXPERIMENTAL VARIABLE (34 *S*s PER CONDITION)

at $p < .0005$. A similar test indicates that the Satiation condition of this study did not reduce the reinforcing effectiveness of approval to a zero level; the mean reinforcer effectiveness score following Satiation (.13) was reliably greater than zero at $p < .01$. Hence, it may be concluded that approval was effective as a reinforcer under all conditions, and that satiation was not complete but only relative.

Spontaneous Social Initiations

An examination of *S*s' spontaneous verbal initiations to *E* during *only* the four-minute baseline period of the game can reinforce the conclusions drawn from the reinforcer effectiveness score (which reflects only the susceptibility of *S*s to approval as a reinforcer). These initiations could be expected to represent behaviors by *S* for a variety of social reinforcers from *E* and, hence, might be sensitive as well to the treatments. It should be noted that *E* did not dispense approval during the baseline period, but merely looked on as *S* put marbles into the toy. (Because the Satiation treatment consisted of a period of intensive social contact between *E* and *S*, which could encourage *S* to continue to make social initiations during the game, only a comparison of the verbal social response scores between the Deprivation and Nondeprivation groups would be meaningful.)

The *total frequency of verbal initiation* scores was classified in a 2 × 2 factorial (Deprivation vs. Nondeprivation and Boys vs. Girls) and subjected to an analysis of variance. The mean number of verbal responses (Comments, Questions, and Attention-seeking) after Deprivation (4.4) was found to be reliably greater than that after Nondeprivation (2.3) at $p < .05$ ($F = 3.70$, 1 and 64 *df*, one-tail test). At the same time, the Treatment × Sex interaction effect was not reliable, indicating that the two treatments did not affect the sexes differentially.

The scale scores represent what appears to be the *intensity of social contact*, or the degree to which certain social reinforcers apart from approval (e.g., attention) are sought from the adult, and could take any one of the four values from zero to three. They were classified in a fourfold table, with *low* scores of zero and one taken to represent less intense social contact, and *high* scores of two and three taken to represent more intense social contact. It was found that 20 out of the 34 *S*s exhibited high scale scores after Deprivation, while only 12 of the 34 *S*s exhibited such high scale scores after Nondeprivation, indicating at $p < .05$ (chi square corrected for discontinuity = 2.89, 1 *df*, one-tail test) that a greater proportion of *S*s exhibited more intense social contact scores after Deprivation than after Nondeprivation.

Hence, the two relatively independent indices derived from the three categories of spontaneous verbal initiations to *E* *both* reinforce the conclusions derived from the measure of the reinforcing effectiveness of *E*'s approval.

DISCUSSION

In the earlier study cited (as in this study), the experimental operation of brief social isolation of the child was equated to a condition of deprivation[6] of all social reinforcers (including approval). Hence, it became important to demonstrate that this condition increases the incidence of behaviors for approval relative to an empirically defined level of satiation for that reinforcer. The results suggest that this has been accomplished. And since under the Nondeprivation condition *S*s had just come from class, with no further experimental treatment, it seemed reasonable that it would represent some point intermediate between the other conditions. The rank-order of the mean reinforcer effectiveness scores for these three conditions followed this logic precisely and was uncomplicated by an interaction involving sex of *S*s. Hence, the results replicate the finding of the earlier study concerning the greater effectiveness of approval after Deprivation, relative to Nondeprivation. While the mean difference between reinforcer effectiveness scores for Deprivation and Nondeprivation for the *S*s in this study (ranging in age from 6-6 to 9-0) appears to be of the same order as that for the *S*s in the earlier study (who ranged in age from 3-10 to 5-3), the absolute level of the means appears to be higher.

---

[6] An analysis of some possible Deprivation condition artifacts is made in (6).

The two indices drawn from the three spontaneous verbal social initiation categories reinforced the conclusions based on the index of the reinforcing effectiveness of *E*'s approval for the Deprivation and Nondeprivation conditions (where a difference test was meaningful). It seems likely then, if the game-set were dispensed with and the natural or spontaneous social initiations to *E* were employed as the sole dependent variable, that similar conclusions would be drawn from the results. However, if the present establishing operation for satiation were to be employed in such a study, it would still be almost impossible to separate the social effects of satiation from the artifactual effects of *S*s' greater familiarity with *E*.

The essence of the parallel between our use of the term drive for social reinforcers and the traditional use of the term for the deprivation of food and water reinforcers, as in hunger and thirst, lies in the definition of a drive as the functional relation between deprivation for a reinforcer and responding for that reinforcer. In that special sense, then, it would appear that there exist for children social drives that respond to social reinforcer deprivation similarly as do many primary appetitive drives. Yet it should not be supposed that the results of this study validate decisively the conclusion that a social reinforcer follows a deprivation-satiation logic. The responsiveness of a reinforcer to relative values of both deprivation and satiation represents only one requirement under this logic. Social reinforcers may be supplied and deprived in a variety of ways, and it is important to discover their responsiveness to many of these ways. It would be especially important, for example, to implement the deprivation of a *single* social reinforcer, rather than of all social reinforcers. Further, it is essential to have some assurance that the social reinforcers are more or less homogeneous in this regard, for verbal approval may not be representative of the reinforcers controlling the purely social initiations made by children to adults. Replications of these effects with other social reinforcers (e.g., attention) would strengthen the conclusions, as would parametric studies of the deprivation-satiation dimension. Another assumption is that approval had acquired reinforcing value for children through a history of conditioning. A stronger case would result if approval were demonstrated to be a more effective reinforcer than, say, a comparable nonsocial noise produced by a machine; or if that noise were shown to be less affected than approval by deprivation-satiation conditions like those employed here. Even so, the earlier finding (6) that isolation enhanced the reinforcing power of adult approval for *S*s as a positive function of the degree to which they sought adult approval in other settings would indicate against both these possibilities. And the difference in spontaneous social initiations following Deprivation and Nondeprivation supports the more general conclusions drawn here; those behaviors were very likely employed for social reinforcers other than approval.

While the results of this study can stand in their own right, several additional experimental conditions could help elucidate the processes at issue. Thus, it would be useful to separate the effects of some of the three components of the satiation condition employed, namely, (*a*) *E*'s physical presence, (*b*) the social interaction between *E* and *S*, and (*c*) the approval reinforcers dispensed by *E* to *S*. To do this, it would be necessary to employ such conditions as, for example, one in which *E* sits with *S* for 20 minutes but says nothing, or one in which approval is dispensed without *E*, possibly out of "thin air." While some such conditions might be difficult to implement, the attempt to establish conditions like these would be worth while in the context of this experiment. Even so, the results of an experiment (3) on related behavior (for attention) suggest that a condition in which *E* sits near *S* but interacts minimally with him (only upon request) functions more as a condition of social deprivation than as one of satiation (7).

Apart from these considerations, the standard primary-conditioned reinforcer issue in the study of animal drives cannot be involved as such in this case. We have assumed that some sort of conditioned reinforcer is at issue, but haven't separated out conclusively the effects of the several stimulus components of that conditioned reinforcer. At the very worst, then, it isn't clear that the social reinforcer dispensed is independent of *E*. We have already noted the possibility that noise, rather than social noise, might constitute the reinforcing stimulus. Apart from this, however, it does not seem significant to consider social reinforcers as separate from the person dispensing them: they are eminently social noises in our logic; and their social character refers not only to their presumed history but also to the method of their delivery.

Additional considerations. In the earlier study, two *E*s were employed to dispense the social reinforcer, one male and one female. An interaction effect was found which indicated, among other things, that the increase in the effectiveness of approval as a reinforcer, brought on by Deprivation relative to Nondeprivation, was greatest for boys with the female *E*. This change was not reliable for the other three Sex of *S* × Sex of *E* groups. In this experiment, in which both boys and girls were tested by a female *E*, no Treatment × Sex interaction was found (although the Satiation mean of girls may not have differed from zero). Hence, to that extent, the previous results are not replicated, but the conclusions can be drawn more generally. (This finding could be referred to the postulate that *S*s of this study are "latent" [in the Freudian sense], while those in the earlier study were "Oedipal." However, there exist other possibilities, including slight changes in method, to explain the discordant results of the two studies on this issue.)

Only one reinforcer-dispensing *E* was employed, and she was not unaware of the expected direction of the results. Her behavior was scored by objective and easily discriminated criteria and judged satisfactory. Still, there remains some reason for caution; for the findings of this experiment cannot be conclusive until they are widely replicated with similar and improved procedures. Nevertheless, the reliability of the results is supported by the difference between the Deprivation and Nondeprivation groups on the spontaneous social initiations made to *E* before she began to dispense the reinforcer: there, where it was even less likely that *E* could influence the outcome, the finding paralleled precisely that derived from the reinforcer effectiveness index.

SUMMARY

On the assumption that approval is representative of the reinforcers controlling the purely social initiations of children to adults, the verbal approval of a female *E* was made contingent upon one response in a two-response game for 102 first- and second-grade children (*S*s) of a university laboratory school. The change in relative response frequency from a baseline level following introduction of the reinforcer indicated the degree to which approval was reinforcing for an *S*.

Before playing the game, *S*s were subjected to one of three experimental conditions: 34 *S*s were subjected to a 20-minute period of isolation (conceived to be social *Deprivation*); 34 *S*s played the game immediately upon coming from class (conceived to be an intermediate condition between deprivation and satiation, and called *Nondeprivation*); and 34 *S*s devoted 20 minutes to drawing and cutting out designs, while *E* maintained a stream of friendly conversation with each of them and approved and admired their art efforts and statements in an appropriate fashion (conceived to be *Satiation* for approval and social contact). There were 17 boys and 17 girls in each experimental group.

Employing two independent tests (one on the order of the means, the other on the regression of the means on the treatment dimension), it was found that the rank-order hypothesis advanced was reliably supported (at $p < .03$ and $p < .005$ respectively). The reinforcing effectiveness of approval was relatively greatest after Deprivation, intermediate after Nondeprivation, and least after Satiation. Approval functioned as a reinforcer after all conditions, even Satiation ($p < .01$). Boys and girls were affected similarly on this measure by the experimental conditions.

The spontaneous social initiations made by *S*s to *E*, before she had begun to dispense approval in the game, supported the results based on the index of the effectiveness of approval as a reinforcer. Following the Deprivation condition, there resulted a greater mean frequency of social initiations ($p < .05$), and a larger proportion of *S*s exhibiting apparently intense social interaction ($p < .05$), than there did following the Nondeprivation condition. (The nature of the establishing operation for the Satiation condition precluded a meaning-

ful comparison of the other conditions with it on these social behaviors.)

Thus, a reinforcer appearing to be typical of those involved in children's social drives appears responsive to deprivation and satiation operations of a similar order as those controlling the effectiveness of reinforcers of a number of the primary appetitive drives.

REFERENCES

1. Bowlby, J. Maternal care and mental health. *Bull. World Hlth. Org. Monogr.*, 1951, **3**, 355–533.
2. Chase, Lucille. Motivation of young children: An experimental study of the influences of certain types of external incentives upon the performance of a task. *Univer. Iowa Stud. Child Welf.*, 1932, **5**, No. 3. Pp. 119.
3. Gewirtz, J. L. Three determinants of attention-seeking in young children. *Monogr. Soc. Res. Child Develpm.*, 1954, **19**, No. 2 (Serial No. 59).
4. Gewirtz, J. L. A factor analysis of some attention-seeking behaviors of young children. *Child Develpm.*, 1956, **27**, 17–36.
5. Gewirtz, J. L. A program of research on the dimensions and antecedents of emotional dependence. *Child Develpm.*, 1956, **27**, 205–221.
6. Gewirtz, J. L., & Baer, D. M. The effect of brief social deprivation on behaviors for a social reinforcer. *J. abnorm. soc. Psychol.*, 1958, **56**, 49–56.
7. Gewirtz, J. L., Baer, D. M., & Roth, Chaya H. A note on the similar effects of low social availability of an adult and brief social deprivation on young children's behavior. *Child Develpm.*, 1958, **29**, 149–152.
8. Goldfarb, W. Psychological privation in infancy and subsequent adjustment. *Amer. J. Orthopsychiat.*, 1945, **15**, 247–255.
9. Greenspoon, J. The reinforcing effects of two spoken words on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.
10. Guttman, L. The basis of scalogram analysis. In S. A. Stouffer *et al. Measurement and prediction.* Princeton, N.J.: Princeton Univer. Press, 1950. Pp. 60–90.
11. Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
12. Hurlock, Elizabeth B. The value of praise and reproof as incentives for children. *Arch. Psychol., N.Y.*, 1924, **11**, No. 71, 5–78.
13. Ribble, Margaret. Infantile experience in relation to personality development. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944. Pp. 621–651.
14. Skinner, B. F. *Science and human behavior.* New York: Macmillan, 1953.
15. Spitz, R. A. Hospitalism: An inquiry into the genesis of psychiatric conditions in early childhood. In Anna Freud *et al.* (Eds.), *The psychoanalytic study of the child.* Vol. 1. New York: Int. Universities Press, 1945. Pp. 53–74.
16. Wolf, Theta H. The effect of praise and competition on the persistent behavior of kindergarten children. *Univer. Minn. Child Welf. Monogr. Series*, 1938, No. 15. Pp. 138.

Received July 12, 1957.

THE EFFECTS OF LEADERSHIP STYLE AND LEADERSHIP POWER UPON THE INDUCEMENT OF AN ATTITUDE CHANGE[1]

DAVID KIPNIS

U. S. Naval Personnel Research Field Activity, Washington, D. C.

Much has been written on the role of participatory leadership in inducing changes in behavior. The early findings of Lewin (7) and Coch and French (3) indicated that when a leader used a participatory style of leadership, as contrasted with some less group-centered style, individuals were more likely to change their behavior in the direction advocated by the leader. More recent studies have replicated these findings (6, 10). However, attempts to demonstrate the usefulness of participatory leadership in areas other than the induction of behavior have produced mixed results. The findings of Torrance (11) on problem solving and Berkowitz (1) on group members' satisfaction with a conference leader, for instance, indicate that participatory leadership is not necessarily superior. While it would appear that participatory leadership may be more effective than other styles under certain conditions, these conditions have not been fully identified.

One condition that may affect the success of participatory leadership is the kind of social power held by the leader. By social power is meant the accorded right of an individual to reward compliance or punish noncompliance by conferring or withholding objectives desired by followers. In many instances, a leader is endowed with such power by virtue of his role in a social structure. Quite apart from the style with which a person exercises leadership, it is possible that the perception of the leader's power as either rewarding or punitive is an important determinant of his influence. That is, influence acts originating from a power source may be more readily complied with when the source is perceived by the group member as helping him towards some desired region than when the power source is perceived as blocking access to a desired region or coercing the individual to remain in, or to enter, some undesired region.

The present study investigates the relative efficacy of participatory leadership and lecture leadership in inducing an attitude change when the leader is perceived as having power to reward compliance or to punish noncompliance.

METHOD

The procedure was adapted from Kelman (5). He showed that when given a choice, most children favor fantasy hero comics (Superman, Batman, etc.) over jungle comic books (Tarzan, Jungle Jim, etc.). In the present study, an attempt was made by a leader to induce children to give up their preference for fantasy hero comics in favor of jungle comics. Since a preference for fantasy hero comics tends to be strongly held, success in inducing an attitude change was measured by the extent to which subjects (*S*s) accepted the view that fantasy hero comics were "bad reading" for children younger than they were.

Subjects and Groups

The *S*s were pupils in nine classes in the 5th and 6th grades of a large public school in one of the poorer areas of New York City[2] (N = 195; 88 boys, 107 girls). Their average age was 11 years. *S*s were randomly assigned to 35 groups according to sex; 16 groups were composed of boys and 19 of girls. The size of the groups ranged from four to eight, with an average of six members. These 35 groups were assigned equally among the six experimental conditions except for one condition (Lecture-Punishment) which contained five groups.

Attitude Ratings

Two folders (A and B) were prepared, each containing the first two pages of 12 comic books. The comics in both forms contained the same "heroes," but in differing adventures. Six of the comics in each folder were fantasy hero comics and six were jungle comics. Inserted after each comic was the following rating scale:

---

[1] The present work is based upon a dissertation submitted to the Graduate School of Arts and Science of New York University in partial fulfillment of the requirements for the degree of doctor of philosophy. The writer wishes to acknowledge with gratitude the encouragement and help of James N. Farr and the many valuable suggestions of Isidor Chein. The cheerful presence and support of his wife, Dorothy M. Kipnis, is also gratefully acknowledged.

[2] It is the writer's impression that the threat of punishment induced less direct and indirect verbal attacks upon the leader by group members during the main study than during a pilot study conducted among upper middle-class children. In the main study, the threat of punishment was usually greeted by silence and a slight withdrawal of interest.

I think this kind of story is: (check one)
( ) Very good reading for younger children
( ) Good reading for younger children
( ) Neither good nor bad reading for younger children
( ) Bad reading for younger children
( ) Very bad reading for younger children

One week prior to the procedures intended to influence attitudes, the folders were distributed by teachers to the participating classes, one half of each class receiving Form A and the other half Form B. In order not to identify these ratings with the experimental procedures to be conducted the next week, pupils were told that the ratings were being done for their school principal who wanted advice on the kinds of comic books they felt were good reading for younger children.

Experimental Conditions

Attempts were made to establish six experimental conditions as follows:

1. Participation—leader had power to reward compliance (Participation–Reward).
2. Participation—leader had power to punish noncompliance (Participation–Punishment).
3. Participation—leader had no power to reward or punish (Participation–Control).
4. Lecture—leader had power to reward compliance (Lecture–Reward).
5. Lecture—leader had power to punish noncompliance (Lecture–Punishment).
6. Lecture—leader had no power to reward or punish (Lecture–Control).

Two experimenters (*E*1 and *E*2) conducted the experimental meetings and alternated in the role of leader. *E*2 (or *E*1) would meet with a group and inform them that he was a professor from a local university and was meeting with them to take a public opinion poll. He would then read the questions upon which they were to vote:

1. Which kinds of stories are better reading (check one)?
( ) Fantasy hero comics
( ) Jungle comics
( ) Both kinds of comics

2. Which kinds of stories are better reading for younger children?
( ) Fantasy hero comics
( ) Jungle comics
( ) Both kinds of comics

3. Why do you like the kind of comics you voted for? ______________________

At this point, *E*2 would introduce *E*1 as a representative of the Jungle Comic Book Publishing Company and say that the school principal had given the publisher permission to address the group. The publisher would tell the group he was there because he had heard that most of them, in the ratings of comic books for their principal, had voted against jungle comic books. When he heard that the college professor was coming to their school to conduct an opinion poll, he had asked permission to talk to them on why jungle comics were also good reading.

The experimental variables were then introduced. The first step established differences in leadership incentives:

Power to Punish. The publisher said that he had been given free movie passes for the local movie which were to be distributed to the entire 5th and 6th grades. Unfortunately, he had discovered that there were 30 movie passes short and therefore 30 children would not receive passes. Because it was his responsibility to decide who could not have any passes, he had decided to take away passes from among those in the various groups who did not vote for jungle comics on Questions 1 and 2 of the opinion poll. Of course not all pupils voting against jungle comics would have their passes taken away from them, but the 30 would be chosen from this group. *S*s were told that movie passes would be given out in three weeks.

Power to Reward. The publisher said that he had a very special announcement to make before talking about jungle comics. His company had decided to have a contest among students taking part in the opinion poll. The prizes would be movie passes to the local movie (number of passes equaled $\frac{2}{3}$ size of group). The contest was open to all students who were going to vote for jungle comics on both questions of the opinion poll and the winners would be chosen from those writing the best reasons for reading jungle comics. Prizes would be awarded in three weeks after the votes had been checked and reasons for liking jungle comics were judged.

Control Conditions. The publisher did not offer a reward for compliance or threaten to punish noncompliance. He merely began the discussion of jungle comics. At the end of the meeting, *E*2 announced he was giving all *S*s free movie passes for their cooperation. These would be given out in three weeks.

The next step in the procedure established differences in leadership style:

Participation. The publisher rapidly outlined four points in favor of jungle comics or against fantasy hero comics and then invited a general discussion of his points. In all cases, the publisher's role was nondirective, and he encouraged as much expression of opinion from the group as possible. Criticisms of jungle comics were accepted and other members of the group were asked to answer if they so chose. After 10 to 12 minutes, *E*2 would interrupt the discussion. Before leaving the room, *E*1 would sum up the meeting by saying that he gathered that most of the group liked jungle comics much better than before.

Lecture. The publisher told the group that because he had only 10 minutes he would not be able to allow anyone in the group to talk or comment on what he was about to say. He then delivered a lecture elaborating the four points favoring jungle comics. No discussion was allowed. At the end of 10 minutes, the lecture was stopped and the publisher expressed the hope that the group liked jungle comics better than before.

Postinfluence Measures

After the publisher left the room, *E*2 assured the group that he did not care which kind of comic book they voted for. He then distributed the opinion forms and allowed *S*s to vote. To obtain a measure of *S*s' perceptions of the experimental variables, a questionnaire was given to the groups following the vote. The groups were assured that while the publisher (called comic book man in the questionnaire) would see the voting form, he would not be allowed to see the questionnaire.

Degree of Private Acceptance

One week after the influence procedures, a second set of comic book folders was distributed by teachers to their classes. *S*s who initially rated Form A now rated Form B, and those who rated Form B, now rated Form A. A statistical measure of *S*s' private acceptance of the leader's influence attempt was obtained as follows from changes from preinfluence to postinfluence attitude ratings of jungle and fantasy hero comics:

1. For the preinfluence attitude ratings, the sum of the ratings for the six fantasy hero comics and the six jungle comics were obtained. Total scores for fantasy hero comics were subtracted from total scores for jungle comics. A negative difference score (Diff 1 score) indicated that *S*s initially favored fantasy hero comics. On the average, *S*s did initially favor fantasy hero comics (Mean Diff 1 score, Form A = −1.22; Form B = −2.08, CR A vs. B = 1.51).
2. For the postinfluence attitude ratings, difference scores (Diff 2 scores) were once again computed.
3. Changes in ratings from preinfluence to postinfluence attitude ratings were computed by subtracting Diff 2 scores from Diff 1 scores. Change scores could range from +48 to −48 points; positive scores indicated shifts in favor of jungle comics.

The average Change score for all *S*s participating in the experimental meetings was +3.65. This change is significantly greater than zero. Fifteen *S*s completed both the preinfluence and postinfluence attitude ratings but were absent from the experimental meetings. Their average Change score was −1.47, not significantly different from zero. The difference between the mean Change score of *S*s participating and not participating in the experimental meetings is significant beyond the .01 level (CR = 6.17).

RESULTS

Effectiveness of the Experimental Variables

Analyses of *S*s' responses to the questionnaire items concerned with participation and perception of leadership power were performed to determine whether the experimental manipulations produced the desired results (see Table 1).

Four questions were included on participations.[3] Two items were directly concerned with whether *S*s were allowed to participate. The remaining items were included under the assumption that one function of participation is to allow members to perceive the opinions of their peers. Weights were assigned to the alternatives of each question and summed over all four items in a single Index of Participation with a range of from 0 to 10. High scores indicated a high degree of participation. From Table 1 it may be seen that mean Participation Index scores for *S*s under participatory conditions are higher than for *S*s under lecture conditions. Analysis of variance of these data indicate that the differences between the two leadership styles are significant beyond the .01 level.

TABLE 1
EFFECTIVENESS OF THE EXPERIMENTAL MANIPULATIONS
(Questionnaire Responses)

Leadership Incentives	Leadership Style[a]		Source[b]	*df*	*MS*	*F*
	Participation Mean	Lecture Mean				
Index of Participation						
Reward	6.68	4.83	Leadership	1	31.93	48.52**
Control	6.35	5.08	Incentives	2	.70	
Punish	7.09	4.43	L × I	2	.71	
			Error	29	.658	
Index of Leader's Power						
Reward	7.05	7.10	Leadership	1	.10	
Control	6.24	6.00	Incentives	2	5.17	6.44**
Punish	5.87	5.77	L × I	2	.05	
			Error	29	.803	

[a] The average score of each experimental group is used as raw score data.

[b] To correct for the fact that the Lecture-Punishment condition has only 5 groups, the mean of these 5 groups is used as the sixth score. One degree of freedom has been subtracted from total degrees of freedom in evaluating the results. This is a correction adapted from one suggested by Lindquist (8, p. 148).

** Significant beyond the .01 level of confidence.

The second experimental variable was the association of the leader with power to reward compliance or to punish noncompliance. Two questions were included to test the effects of this variable, and the responses to these questions were also combined into a single score.[4] This resulting Index of Leader's

---

[3] Questions were: 1. Who did the most talking? 2. Did you know how the other kids were going to vote? 3. Did the comic book man let you talk? 4. Which kind of comic books did the kids really like?

[4] Questions were: 1. Will the comic book man punish some kids not voting for jungle comics? 2. Will the comic book man give a special prize to some kids who vote for jungle comics? Reverse weights were given to Q. 1 so that a "Yes" answer was weighted zero. For Q. 2, a "No" answer was weighted zero. In the Index, scores in the middle range indicate the leader was perceived as neither rewarding nor punitive.

Power had a range of from 2 to 8, with low scores indicating that the leader was perceived as holding punitive power. Table 1 shows that the differences between incentive levels are significant beyond the .01 level and that the order of mean values is in the expected direction.

Public Compliance

Public compliance was defined as voting for jungle comics on Question 1 and Question 2 of the opinion poll. The percentages of public compliers under each experimental condition are given in Table 2. In general, the percentages of public compliers are larger under conditions of reward and punishment than under the control conditions.

A chi-square analysis of these data indicates no significant over-all differences in the percentages of public compliers in each condition (chi square = 6.87, *df* 5). In view of the lack of significant over-all effects, the results of a more detailed analysis are only suggestive. It may be noted, however, that both reward and punishment produced more public compliance than the control conditions. Approximately 57 per cent of *S*s under conditions of reward complied, 59 per cent under punishment, and 40 per cent under control conditions. Differences between combined reward and combined control as well as differences between combined punishment and combined control are significant beyond the .05 level. These results are similar to those of a joint study by McBride (9) and Burdick (2), who found that both reward and punishment induced significantly more public compliance than a control condition, but that reward and punishment did not differ significantly from each other.

TABLE 2
PERCENTAGES OF PUBLIC COMPLIERS AND NONCOMPLIERS

Condition	Public Compliers %	Noncompliers %	Total *N*
Participation-Reward	61.3	38.7	31
Participation-Punish	54.3	45.7	35
Participation-Control	40.6	59.4	32
Lecture-Reward	54.5	45.5	33
Lecture-Punish	63.6	36.4	33
Lecture-Control	38.7	61.3	31

Chi square = 6.87, *df* 5, *p* ns.

Private Acceptance in Relation to Experimental Conditions

For purposes of classification, Change scores of $+2$ or higher were considered to reflect favorable change in private beliefs; Change scores of $+1$ or less were taken as indicating either no change in the *S*'s original views or a change more favorable toward fantasy hero comics. Table 3 gives the percentage of *S*s within each experimental condition who changed to a more favorable attitude toward jungle comics. Using Wilson's analogue of analysis of variance (13), significant effects were found only for the interaction of Leadership Style and Leadership Incentives (chi square L × I = 15.36, *df* 2, $p < .01$). The significance of the differences between leadership styles at each incentive level are also given in Table 3.

It is seen in Table 3 that participatory leadership induced significantly more *S*s to change their beliefs when the leader was associated with neutral power or power to reward compliance. However, when the leader threatened to punish noncompliance, there was a tendency ($.10 > p > .05$) for lecture leadership to induce *more* *S*s to change their beliefs in favor of jungle comics than changed under participatory leadership.

Since the chi-square value for the interaction of the two experimental variables was significant, the statistical significance of the differences in the number of changers within each leadership style was also tested. These findings are also given in Table 3, which shows that the threat of punishment induced differing reactions among *S*s depending upon whether the leader adopted a participatory or lecture style of leadership. Under participatory leadership, the threat of punishment induced significantly fewer *S*s to privately change their beliefs than changed under Participation-Control conditions. Under lecture leadership the reverse was true.

Neither Participation-Reward nor Lecture-Reward conditions contain significantly more changers than were found in their respective control conditions. While this lack of significant results was contrary to expectations, it is

TABLE 3

PERCENTAGE OF *S*s CHANGING THEIR BELIEFS IN FAVOR OF JUNGLE COMICS

Leadership Incentives	Leadership Style Participation %	*N*	Lecture %	*N*
Reward	84	(31)	51	(33)
Control	78	(32)	39	(31)
Punish	49	(35)	70	(33)

Comparison	*CR*
Between Leadership Styles	
Participation-Reward vs. Lecture-Reward	3.05**
Participation-Control vs. Lecture-Control	3.42**
Participation-Punish vs. Lecture-Punish	1.81*
Within Leadership Styles	
Participation-Reward vs. Participation-Control	.61
Participation-Reward vs. Participation-Punish	3.30**
Participation-Control vs. Participation-Punish	2.61**
Lecture-Reward vs. Lecture-Control	.88
Lecture-Reward vs. Lecture-Punish	1.71*
Lecture-Control vs. Lecture-Punish	2.63**

* Significant between the .05 and .10 level.
** Significant beyond the .01 level.

possible that the lack of significant differences resulted from the experimental procedures used to offer the reward. Some *S*s had expressed doubts about their ability to write compositions. While these *S*s were reassured, it may be that some of them did not attempt to compete for the reward.[5]

Findings for Ss *for whom Manipulations were Effective*

Not all *S*s indicated on the questionnaire that they perceived the experimental setting as structured by the experiment. For purposes of further analysis, *S*s were discarded whose scores on the questionnaire indices of Participation and Leadership Power deviated from expected scores for their condition. A total of 85 *S*s were eliminated, leaving 15 *S*s for Participation-Reward, 15 for Participation-Control, 15 for Participation-Punishment, 24 for Lecture-Reward, 19 for Lecture-Control, and 22 for Lecture-Punishment.

TABLE 4

PERCENTAGES OF *S*s CHANGING THEIR BELIEFS IN FAVOR OF JUNGLE COMICS

(Only *S*s responding to the experimental variables)

Leadership Incentives	Leadership Style Participation %	(*N*)	Lecture %	(*N*)	Comparison	*CR*
Reward	87	(15)	54	(24)	Participation-Reward vs. Lecture-Reward	2.48*
Control	80	(15)	42	(19)	Participation-Control vs. Lecture-Control	2.58**
Punish	47	(15)	86	(22)	Participation-Punish vs. Lecture-Punish	2.61**

* Significant beyond the .05 level.
** Significant beyond the .01 level.

The analysis of Change scores in this purified group intensifies the results obtained using all cases (Table 4). The difference between the number of *S*s changing their beliefs under Participation-Punishment and Lecture-Punishment conditions rises from the .08 level to the .01 level of confidence. Thus the trend that was apparent in the main analysis, that fewer *S*s changed beliefs under Participation-Punishment than under Lecture-Punishment, receives additional support. Comparisons of the differences within each leadership style are not presented here since the results duplicate those using all cases with one minor exception.[6]

Affect Associated with the Leader

It was assumed that the perception of the leader's power as either rewarding or punitive would be accompanied by differences in the affective evaluations of the leader. Three questions were included in the questionnaire to test this assumption.[7] These questions were combined to give an Index of Affective Response. Its range was from 3 to 9, with higher scores indicating a favorable perception of the

---

[5] A comparison of Tables 2 and 3 shows that the percentages of private compliers are greater than the percentages of public compliers. These percentages are not directly comparable, since the use of a different cutting score for defining private compliers would have resulted in differing percentages. Further, the instruments as well as the questions on which *S*s voted differed for the two indices.

[6] Using the purified cases, the comparison between Lecture-Punishment and Lecture-Reward was significant beyond the .05 level (CR = 2.48).

[7] Questions were: 1. Was the comic book man fair? 2. Would you like to meet with the comic book man again? 3. Was the comic book man a good guy?

TABLE 5

INDEX OF AFFECTIVE RESPONSE TO THE LEADER

(Only *S*s responding to the experimental variables)

Leadership Incentives	Leadership Style[a]				Source[b]	*df*	*MS*	*F*
	Participation		Lecture					
	Mean	(*N*)	Mean	(*N*)				
Reward	8.79	(15)	7.58	(24)	Leadership	1	.85	5.84*
Control	8.47	(15)	7.00	(19)	Incentives	2	1.00	6.88**
Punish	6.67	(15)	7.09	(22)	L × I	2	.405	2.78
					Error	104	.145	

[a] Individual scores are used as raw score data.

[b] Because cells contain unequal *N*s, a correction suggested by Walker and Lev (12, p. 381) was used.

* Significant beyond the .05 level.

** Significant beyond the .01 level.

leader. Table 5 presents the average rating of the leader on this index and the results of an analysis of variance of these data. To conserve space, only the findings based upon the purified sample are given.

Affective evaluations of the leader varied with both the type of power he held and his style of leadership. In general, *S*s under participatory leadership evaluated the leader higher than *S*s under lecture leadership. Further, *S*s under reward conditions evaluated the leader higher than *S*s under punishment conditions.

Since the variance due to leadership style was significant, a comparison of individual means was carried out to determine whether participatory leadership induced greater liking for the leader at each incentive level. Similarly, since the variance due to leadership incentives was also significant, means for style of leadership were combined and the significance of the differences between incentive levels were computed. The results of these tests of significance are given in Table 6.

TABLE 6

INDEX OF AFFECTIVE RESPONSE TOWARD THE LEADER

(Critical ratios of differences between leadership style and between incentives[a])

Condition	*df*	*CR*
Between leadership styles		
Participation-Reward vs. Lecture-Reward	104	2.16*
Participation-Control vs. Lecture-Control	104	2.56*
Participation-Punish vs. Lecture-Punish	104	0.59
Between incentive levels		
Reward vs. Control	104	1.15
Reward vs. Punish	104	3.38**
Control vs. Punish	104	2.15*

[a] The error estimate for comparing individual means is based upon the mean square error term obtained in the analysis of variance (8, p. 90).

* Significant beyond the .05 level.

** Significant beyond the .01 level.

It is seen in Table 6 that participatory leadership induced significantly higher affective evaluations of the leader than did lecture leadership under conditions of reward and control. This finding supports the belief that participatory leadership has "human relations" value in reducing barriers between leaders and subordinates. However, it should be noted that no differences were found between participatory and lecture leadership in affective evaluations of the leader under conditions of punishment, and *S*s under punishment expressed significantly less liking for the leader than under reward or control conditions.

DISCUSSION

In most studies concerned with overcoming resistance to change, the norms governing the behavior being changed are rooted in a reference group to which the leader does not belong. It is fairly well agreed that one important means of effecting change, in these instances, is for the leader to allow individuals to perceive the course of action of other members of the normative group. Group discussion is assumed to be effective here because it facilitates communication between group members and thus allows them to perceive that others are ready to change. To a great extent the group rather than the leader acts as the inducing agent. The function of the leader, using group discussion, is to redirect group forces in a direction congruent with his own.

Participatory leadership has sometimes been viewed as an automatic means of redirecting group forces in a direction congruent with the leader's. The present results suggest that things are more complicated. While participatory leadership was more effective than lecture leadership in inducing private acceptance under reward and control conditions, participatory leadership under punishment conditions resulted in *fewer* *S*s privately accepting the leader's influence attempt than under Lecture-

Punishment condition or under Participation-Control conditions.

It may be recalled that when they were threatened with punishment for noncompliance, *S*s under both styles of leadership expressed low affective evaluations of the leader. This result is not too surprising since the leader was attempting to induce *S*s to change established attitudes under threat of punishment for maintaining them. These low affective evaluations can be attributed to a desire to resist the leader's influence attempt. Why, then, did *S*s under participatory leadership successfully resist the leader's influence while *S*s under lecture leadership did not?

It is plausible to suppose that participatory *S*s in their discussion received support from each other which encouraged them to resist the leader's induction attempt. Under the Lecture-Punishment condition, no support or information was available as to the course of action of fellow group members, and each *S* had to decide for himself whether or not to comply. Group discussion, on the other hand, is assumed to have allowed individual members to perceive both resistance among peers and an unwillingness to change standards. Thus the group discussion under threat of punishment served to reinforce original standards rather than to change them.

Considering next the affective evaluations of the leader, the findings provide cues as to why participatory leadership usually engenders more liking for the leader than does lecture leadership. Participatory leadership generally allows the group the illusion that the leader has relinquished a share of his powers and vested them in the group. It follows that there is less awareness by the group of direct pressure to change originating from the leader. It is suggested that this reduction in direct force from the leader results in more favorable evaluations of him.

It may be seen that participatory leadership is not always successful in producing favorable evaluations of a leader. In the present study, when followers perceived the leader reinforcing his induction attempts with threats, affective evaluations of him were significantly lower than in the other two participatory conditions and no different than evaluations in the Lecture-Punishment condition. These findings would suggest that the awareness of direct pressure was no longer minimized by a participatory style of leadership. That is, the group was aware that the leader held the "reins" and they did not have a free choice. Accordingly, force to change was still seen as originating from the leader and not, as intended, from the group.

A surprising finding of this study was that the Lecture-Punishment condition induced significantly more *S*s to change their beliefs than changed in the other two lecture conditions. The findings appear to fit the theoretical model of Hovland and his co-workers (4) as to when a persuasive communication that arouses mild anxiety is accepted. If it is assumed that the threat of the leader aroused mild anxiety, then compliance or later wishful thoughts about compliance with the leader's recommendation should reduce the anxiety engendered by the threat and accordingly reinforce the complying response.[8] It should be noted that to the extent anxiety reduction did determine this effect, the fact that more private compliance did not occur under the Participation-Punishment condition suggests that the group discussion served to minimize anxiety aroused by the leader's threat.

CONCLUSIONS

The purpose of this study was to determine whether a participatory style of leadership was more effective than a lecture style of leadership in inducing an attitude change when the leader explicitly offered to reward compliance or threatened to punish noncompliance.

Thirty-five groups of four to eight pupils in the fifth and sixth grades of a public school met with a leader who tried to change their preferences for comic books. Six experimental conditions were established: two conditions in which the leader offered a prize of a movie pass to *S*s complying with his induction attempt;

---

[8] Public noncompliers under punishment conditions did not know who among them would have their tickets taken away. For the above explanation to accord with Hovland's theory, it must be assumed that thoughts about the threatened punishment continued to arouse anxiety, and attempts to reduce anxiety took the form of a silent rehearsal of the leader's recommendations and a certain wishful fantasy that they had been followed. By wishful fantasy is meant such thoughts as: "I wonder if I'm going to get a movie ticket. Maybe I should have voted for jungle comics, they're not so bad."

two conditions in which the leader threatened to take away previously promised movie passes from some of the *S*s not complying with his induction attempt; and two control conditions in which no sanctions were associated with the leader. The leader then conducted a meeting attempting to change *S*s' preferences for comic books using either a participatory or a lecture style of leadership. Measures of public compliance and liking for the leader were obtained at the meeting. Measures of *S*s' private acceptance of the leader's influence were obtained from changes between ratings of comic books obtained one week prior to the experimental meetings and one week after these meetings. The latter measures were thought by *S*s to be independent of the experimental meetings.

The results were as follows:

1. Both reward and punishment produced more public compliance than the control conditions.

2. Participatory leadership induced more *S*s to privately change their beliefs than did lecture leadership when the leader was associated with neutral power or power to reward compliance.

3. When the leader threatened to punish noncompliance, significantly fewer *S*s changed their beliefs under participatory leadership than under lecture leadership or under the Participatory-Control condition. Group discussion under threat of punishment from the leader reinforced original standards rather than changed them.

4. Participatory leadership engendered greater liking for the leader than did lecture leadership as long as the leader was not perceived as enforcing his induction attempt with threats. When threatened with punishment, *S*s under participatory and lecture leadership rated the leader equally low.

REFERENCES

1. BERKOWITZ, L. Sharing leadership in small decision making groups. *J. abnorm. soc. Psychol.*, 1953, **48**, 231–238.
2. BURDICK, H. The compliant behavior of deviates under conditions of threat. Unpublished doctoral dissertation, Univer. of Minnesota, 1955.
3. COCH, L., & FRENCH, R. P. Overcoming resistance to change. *Hum. Relat.*, 1948, **1**, 512–532.
4. HOVLAND, C. I., JANIS, I., & KELLEY, H. H. *Communication and persuasion.* New Haven: Yale Univer. Press, 1953.
5. KELMAN, H. C. Attitude change as a function of response restriction. *Hum. Relat.*, 1953, **6**, 185–214.
6. LEVINE, J., & BUTLER, J. Lecture vs. group decision in changing behavior. *J. appl. Psychol.*, 1952, **36**, 29–33.
7. LEWIN, K. Studies in group decision. In D. Cartwright & A. Zander (Eds.), *Group dynamics.* New York: Row Peterson, 1953.
8. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
9. MCBRIDE, DOROTHY. The effects of public and private change of opinion on intra-group communication. Unpublished doctoral dissertation, Univer. of Minnesota, 1954.
10. PRESTON, M. G., & HEINTZ, R. K. Effects of participatory vs. supervisory leadership on group judgment. *J. abnorm. soc. Psychol.*, 1949, **44**, 345–355.
11. TORRANCE, E. P. Methods of conducting critiques of group problem solving performance. *J. appl. Psychol.*, 1953, **37**, 394–398.
12. WALKER, H. M., & LEV, J. *Statistical inference.* New York: Henry Holt, 1953.
13. WILSON, K. V. A distribution free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, **53**, 96–101.

Received July 17, 1957.

SOCIAL ROLE AND PATTERNS OF SYMPTOMATIC BEHAVIORS[1, 2]

LESLIE PHILLIPS
Worcester State Hospital, Clark University

AND M. SAM RABINOVITCH
Children's Memorial Hospital, Montreal, Quebec

THE present study is a contribution toward a more conceptually oriented system of psychiatric classification than is currently available. The present conventional categories of diagnosis are based primarily on descriptions of overt signs of psychopathology, and are inherently limited in their theoretical base. For example, the precise character of the current diagnostic syndromes themselves is unclear, for these categories sometimes seem to correspond to various (unknown) "disease" processes; sometimes they represent simply "disorders" of "the psychobiological unit"; and sometimes a patient may be "ill primarily in terms of society and of conformity with the prevailing cultural milieu" (1).

Further, there is an ambiguous basis for diagnosis in that a single disorder carries the connotation of a variety of disparate symptoms. Various cases carrying a given diagnosis may be classified on quite different grounds, e.g., one patient may be categorized as schizophrenic because of an affective disturbance, another because of cognitive impairment. Thus, in terms of their contribution to a diagnosis, the symptoms of a syndrome are interchangeable. Indeed, the bases of classification shift among cognitive, affective, mood, social, and neuropsychological factors, and there has been relatively little concern with the specific nature of these individual signs of disorder. Because of the ambiguities inherent in the current system of psychiatric diagnosis, it appears more meaningful to deal directly with the raw data of psychopathology, i.e., those explicit symptomatic behaviors which bring patients into institutional contact. Therefore, we studied the statistical relationships occurring among various symptomatic reactions, independent of existing nosologies. Guided by these results, we organized these deviant reactions into symptom clusters or categories. The final phase of the study involved a test of the stability of these categories in another group of subjects.

METHOD

The study was based on an examination of case history data on 604 patients. The Worcester State Hospital Psychology Department keeps detailed case history information on all patients referred for psychological appraisal. The files for approximately a five-year period (1948–52) were screened, and all cases with two or more presenting symptoms were used in the present study. A symptom, in this sense, refers to the description of a patient's behavior by a psychiatrist at the time of initial institutional contact, or the description of behavior presented by referring physicians as the main reason for hospitalization. These symptoms largely represent specific actions, thoughts, or somatic reactions that provide a comprehensive (although not exhaustive) list of those observed in a typical state hospital population, a psychiatric outpatient clinic, and in jail cases.[3]

The investigation proceeded by randomly dividing the total sample into two groups of 302 cases each. The first, or pilot, group provided the initial data on which the symptom categories were based. The other group of subjects was reserved for later use in testing the stability of these devised categories.

Extremely deteriorated individuals, obvi-

---

[1] This investigation was supported by the Dementia Praecox Research Project, Worcester State Hospital, and a research grant (M-896) from the National Institute of Mental Health, Public Health Service.

[2] Thanks are extended to the members of the Psychology Department, Worcester State Hospital, for their many constructive criticisms in the writing of this manuscript. We are particularly indebted to John E. Lane, who helped extensively in the organization of this paper.

[3] A description of the population sample used in this study has been deposited with the American Documentation Institute. Order Document No. 5637 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

ously organic cases, and very agitated patients are seldom referred to the psychological service and consequently are not adequately represented in either sample.

Statistical Analysis of Data

Examination of the 604 case records resulted in the tabulation of 46 discrete presenting symptoms. Of these, seven were eliminated from the study because their low frequency of occurrence made them unsuitable for the type of statistical analysis chosen.[4]

The data for the remaining 39 symptoms were independently tabulated and cross-tabulated by IBM procedure. Thus, for each pair of symptoms, the number of patients showing both symptoms, neither symptom, and one of the pair, was determined.

All possible pairings of these 39 symptoms result in 741 different pair-combinations. Each such pair was examined for any tendency for (*a*) the presence of one symptom to be associated with absence of the other and (*b*) the simultaneous presence of both symptoms of the pair.

The inference as to whether any such association obtained between paired symptoms was based upon the application of the chi-square test of independence to each of the 741 possible combinations. The .10 significance level was chosen for all tests. A statistically significant result, therefore, indicates that if any two symptoms were indeed independent, the probability is only .10 that the frequency of joint occurrence (or mutual exclusiveness) found in the sample would obtain.

Definition of Symptom Categories

For the pilot group, 31 (out of the 39) symptoms evidenced either positive or negative[5] relationships at the .10 level of significance. From an analysis of the relationships, it is possible to organize 25 of these symptoms into three categories according to the following criteria:

(*a*) Each symptom appears in only one of these categories;

(*b*) Only positive relations occur among symptoms within each category, whereas no positive relations appear between symptoms in different categories;

(*c*) For a given symptom, negative relationships occur only with symptoms in the other two categories.

While the three major categories appear quite clear-cut, two exceptions appeared. These were the occurrence of a positive relationship between "perplexed" in Category 1 and "homosexuality," which was grouped in Category 2. A second is the occurrence of a positive relation between "bizarre ideas" (Category 1) and "headaches" (Category 3). Classification was decided largely on rational grounds for each of these symptoms. No such decision could be made for "fears of own hostile impulses," which showed a statistical relationship with both "drinking" (Category 2) and "depreciatory ideas against the self" (Category 3). Consequently, it was dropped from further consideration. This was true also for "blackouts" and "euphoria," which demonstrated only negative relations and could not therefore be categorized.

Following the formation on statistical grounds of the three symptom categories listed above, a rational analysis of their organization was undertaken. Thus, the first category was taken to represent a withdrawal from others either directly (suspiciousness, withdrawn) or in distorted fantasy (e.g., hallucinations, sexual preoccupation, bizarre ideas). Because "amnesia" may occur on an organic basis, it was deleted from Category 1 and was not further considered. The second category comprises a group of symptoms which were characterized either as actions or thoughts turned against others (e.g., threatens assault, temper outbursts, robbery) or as self-indulgent and socially disapproved behavior patterns (e.g., drinking, perversions). The last of the three categories was interpreted as a turning against the self, either in directly expressive action or thoughts (e.g., suicidal attempt, depreciatory ideas against self) or in suffering expressed somatically (e.g., bodily complaints, insomnia). In essence, this last category appears to represent a group of behaviors which are the opposite of those organized within the second one discussed, in which behavior and bodily expression are so largely turned against others.

On the basis of the rational analysis superimposed on the empirical relationship obtained with the pilot group, the number of symptoms considered for further test was increased. They then also included those deviant reactions which logically belonged in each of the three categories, but for which statistically significant relationships had not been obtained.

The following hypothesis was then formulated: The three categories as described above would be confirmed in the test sample. The specific symptoms in each of these three categories are listed in Table 1.

RESULTS

The results obtained with the test sample ($N = 302$) are in agreement with the formulated hypothesis. Positive statistical relationships were found only within each of the symptom categories and negative relationships were established only between symptom categories. The results of this study suggest that for a given individual, if any one of his symptoms falls into one of the categories defined here, his other symptoms (among those listed above) are also likely to fall into that same category. As shown in Table 2, no exceptions

---

[4] A list of the symptoms used in this study and the frequency of their occurrence have been deposited with the American Documentation Institute. Order Document No. 5637 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[5] For ease of exposition, the term "positive relationship" will be used when there is a tendency for the joint occurrence of two symptoms, and the term "negative relationship" when the occurrence of two symptoms tends to be mutually exclusive.

TABLE 1
LIST OF SYMPTOMS OCCURRING IN EACH CATEGORY

	Ideational Symptomatology	Mood and Affect Disturbances	Behavioral Symptoms	Somatic Disorders
Category 1[a] (Avoidance of others)	withdrawn feels perverted suspiciousness hallucinations bizarre ideas sexual preoccupations	apathetic perplexed		
Category 2 (Self-indulgence and turning against others	threatens assault		drinking rape (and incest) emotional outbursts irresponsible behavior robbery lying disturbed, destructive perversions (including homosexuality) assaultive	
Category 3 (Self-deprivation and turning against the self)	suicidal ideas depreciatory ideas against self	depressed tense-nervous	suicidal attempt compulsions doesn't eat	bodily complaints insomnia headaches

[a] The symptom of "feels going crazy or mentally sick" would have been included in Category 1, and "accused of murder," "firesetting," "eats a lot," and "all addictions except drinking" would have been included in Category 2, but their incidence in the total sample of 604 *S*s was too low for the type of statistical analysis used.

TABLE 2
ACTUALLY OCCURRING VERSUS PREDICTED RELATIONSHIPS AMONG CATEGORIES

	Within Categories		Between Categories	
	Positive	Negative	Positive	Negative
Category 1	8	0	0	19
Category 2	4	0	0	4
Category 3	13	0	0	0
Total	25	0	0	23

to the predicted placement of symptoms occurs in the test sample. As will be seen, the results are so much in conformity with the hypothesis that a statistical test of significance appears superfluous. In any case, it would be difficult to set up any legitimate test of the significance of these findings since there is a lack of statistical independence among the obtained relationships. Table 3 lists those relationships between symptoms which were actually obtained.

Ideally, one might expect interrelationships among all symptoms within a given category, but this happy state does not obtain. In part, this may be due to the fact that a substantial number of the symptom pairs occur with relatively low frequency. Consequently, this precludes the likelihood of obtaining statistically significant relationships with these items. Thus, while in many instances there is insufficient evidence to infer a relationship, this is not to say that we have demonstrated a lack of such relationships. Further, the specific symptom pairs within a category which show positive relationships differ somewhat between the pilot and test groups, as do some of the negative relationships across categories. For example, in Category 1, "suspiciousness," which was positively associated with "perplexed" in the pilot sample, is not so related in the test group. Similarly, for this same category, in the test group "hallucinations" is positively related to "withdrawn," whereas this relationship does not occur in the pilot study. Nevertheless, such specific incongruities must be interpreted within the context of the general results reported; no negative relationships occur within categories and no positive relationships occur between them.

There are other reasons, however, why a question may be raised as to the clear-cut nature of the obtained results. For example, symptoms assumed to be associated with an observed symptom may have been sought and noted by an interviewer while other symptoms were ignored. Illustratively, it is possible that in an assaultive individual evidence for depression may not have been considered. However, case histories were collected over approximately a five-year period, during which there was a considerable yearly turnover in the psychiatric resident staff at the State hospital, so that there was no opportunity for the biases of a restricted number of individuals to distort the reported case information. Further, the frame of reference at this hospital

TABLE 3
RELATIONS BETWEEN SYMPTOMS OBTAINED IN TEST SAMPLE
($N = 302$)

Symptom	Joint Occurrence With	Mutually Exclusive With
1. suspicious	hallucinations sexual preoccupations feels perverted	drinking rape (and incest) other perversions emotional outbursts robbery depressed suicidal ideas suicidal attempt headaches
2. hallucinations	withdrawn perplexed	emotional outbursts depressed headaches
3. bizarre ideas	perplexed	
4. withdrawn	apathetic sexual preoccupations	threatens assault suicidal attempt
5. apathetic		tense-nervous
6. sexual preoccupations		emotional outbursts depressed suicidal ideas
7. drinking	irresponsible behavior	depreciatory ideas vs. self
8. threatens assault	assaultive emotional outbursts	depressed
9. assaultive	emotional outbursts	depressed depreciatory ideas vs. self
10. doesn't eat		depressed
11. depressed	depreciatory ideas vs. self suicidal ideas suicidal attempt tense-nervous bodily complaints	
12. depreciatory ideas vs. self	suicidal attempt tense-nervous	
13. suicidal ideas	suicidal attempt	
14. compulsions	tense-nervous	
15. tense-nervous	insomnia	
16. bodily complaints	headaches insomnia	

during the period when these records were accumulated was that of descriptive psychiatry, and thus a minimum of restrictions based on particular theoretical formulations was probably imposed on symptom categorization.

There is a corollary consideration that both psychiatric residents and psychological interns, who recorded the data for the psychology department, may have ignored actually appearing symptoms. Finally, because the original data were not collected for research purposes, there were no definite criteria defining symptoms. Consequently, there were undoubtedly individual differences among the clinical staff in terms of the intensity of disturbance required for reporting the appearance of a symptom. These uncertainties as to the reliability of the raw data attest to the possible incompleteness of results, and the findings should be interpreted within this context. Thus, in this study, we have probably not succeeded in isolating "pure" clusters, but the evidence for their reality is so suggestive that it calls for further investigation directed toward their refinement and clarification.

DISCUSSION

Since the hypothesis about categories was confirmed, one can presume that for a given individual, symptoms tend to fall into only one of the three categories defined here. The deviant forms of social interaction expressed in these categories of symptomatic behaviors may be characterized as (*a*) avoidance of others, (*b*) self-indulgence and turning against others, and (*c*) self-deprivation and turning against the self.

The characteristics of interpersonal relationships delineated here may be compared to those outlined by Fromm (3), Horney (4), and Rosenzweig (7). Fromm speaks of social relationships in terms of "symbiotic relatedness," "destructiveness," and "withdrawal." Horney describes the neurotic trends of "moving against," "moving away," and "moving towards" other individuals. The formulations of these writers share with each other and with the present findings the general concepts of "avoidance of others" and of "turning against others." Rosenzweig's classification of "extrapunitive" and "intropunitive" types appears to correspond to the interpersonal reactions described here of "turning against others" and "turning against the self." Neither

Fromm's "symbiotic relatedness" nor Horney's "moving toward others" can be assessed from our present data, because these behavioral characteristics were not sampled in the admitting interviews of the population used in this study. It may be also that the positive orientation in interpersonal relationships implied by these terms is deficient in a psychopathological population.

The categories proposed on the basis of the present findings may also be formulated within the context of role theory, i.e., symptom patterns may represent a particular mode of relating (albeit inappropriately) to another person. Particular patterns of role-taking may be presumed to represent the individual's implicit assumptions as to the nature of his relations with himself and with others. In themselves, the role orientations described here do not represent psychopathological phenomena, but if attitudes are patterned exclusively in such negative terms toward the self or others, this might provide the basis for the emergence of socially deviant thought and behavior (as well as for such accompanying somatic reactions as headaches and insomnia).

The question may be posed as to how one might conceptually organize the role orientations as represented in the categories. Perhaps a fruitful way of thinking about this problem is in terms of the principles of developmental psychology.

It has been postulated that development proceeds from a relatively global state to one of increasing differentiation and hierarchization (8). When this formulation is applied to social relationships, it would suggest a genetically early period in which the boundaries between the self and the world are relatively diffuse; a subsequent period of self-centeredness in which a separation of self and world has occurred but with little recognition of the rights of others; and a final phase in which self identity is maintained, but in which relationships on the basis of reciprocity also take place. In this last stage of development, a recognition and acceptance of the social mores has been internalized.

Within a developmental framework, psychopathology represents a relative dominance of genetically early levels of functioning accompanied by a decreasing emphasis on the higher, more recently developed, genetic characteristics. On this basis, the symptom categories described in this study may be rationally ordered in the following sequence. The category "avoidance of others" is considered most primitive, for the chronic threat which social contact implies for the individual may have prevented an adequate development of self and object constancy (5) with a consequent vague and amorphous perceptual structuring of the world. Analogously, Rapaport (6) considers the motor action of withdrawal as a primitive form of behavior upon which is built later forms of reality testing. Recent work by Bexton, Heron, and Scott (2) has demonstrated that isolation from environmental stimulation is regularly followed, even in normal subjects, by hallucinatory experiences. Their findings are pertinent to this present study in that they suggest the possibility that the appearance of hallucinations as well as of bizarre ideas is a consequence of the suspiciousness and withdrawal which characterizes individuals whose deviant behavior falls into the category "avoidance of others."

In a genetic hierarchy of the proposed three classes of symptoms, the category of "self-indulgence and turning against others" appears to fall at an intermediate position. First, the category does not include any symptom indicative of a formal disturbance of thought. Further, the behaviors in this group are suggestive of a self-centered and demanding attitude without concern for the social mores. The salient needs of the individuals whose symptoms fall in this category are of a concrete bodily nature, such as drinking or various deviant forms of sexual expression. From the manner in which these symptoms cluster, it appears that if his needs are not met, the individual is likely to retaliate with behavior carrying an injurious intent. This aggression may be expressed overtly in such behaviors as assault or robbery, or less directly in irresponsible behavior or in threatening assault.

The third category, "self-deprivation and turning against the self," suggests an introjection of social standards not implied by the two previous categories. Thus the individual whose symptoms fall into this last grouping would appear more socially mature and responsible in that he has introjected and can take the role of internalized others. The psychopathology represented by this class of symptoms involves an excessively harsh judgment of the self, leading to self-deprivation (e.g., not eating) or a turning against the self in either action or thought (suicidal attempt, suicidal ideas) or their somatic equivalents (tension, headaches).

SUMMARY

This study has analyzed the tendency of symptomatic behaviors to appear together or be mutually exclusive in the deviant behavior of the individual patient. Thirty-nine symptoms of a sufficiently high frequency of occurrence for statistical analysis were considered. The psychiatric population covered a broad range of diagnostic categories and was drawn largely from a state hospital, although groups from an outpatient clinic and from jail were also liberally represented. A total of 604 patients was studied, distributed randomly into two groups of 302 subjects each. A statistical analysis of the interrelationships of the 39 symptoms under study was undertaken for the first group. On the basis of this exploratory work, it was hypothesized that symptoms tend to occur in three major groupings, which are assumed to be indicative of "avoidance of others," "self-indulgence and turning against others," and "self-deprivation and turning against the self." The validity of these categories was supported by the findings on the second group of subjects. Finally, these categories were interpreted within the context of developmental theory.

REFERENCES

1. AMERICAN PSYCHIATRIC ASSOCIATION. *Diagnostic and statistical manual of mental disorders.* Washington, D. C.: Author, 1952.
2. BEXTON, W. H., HERON, W., & SCOTT, T. H. Effects of decreased variation in the sensory environment. *Canad. J. Psychol.*, 1954, **8**, 70-76.
3. FROMM, E. *Man for himself.* New York: Rinehart, 1947.
4. HORNEY, K. *Our inner conflicts.* New York: Norton, 1945.
5. PIAGET, J. *Psychology of intelligence.* New York: Harcourt, Brace, 1950.
6. RAPAPORT, D. *Organization and pathology of thought: Selected sources.* New York: Columbia Univer. Press, 1951.
7. ROSENZWEIG, S. An outline of frustration theory. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*, Vol. 1. New York: Ronald, 1945.
8. WERNER, H. The concept of development from a comparative and organismic point of view. In D. B. Harris (Ed.), *The concept of development: An issue in the study of human behavior.* Minneapolis: Univer. of Minnesota Press, 1957.

Received July 8, 1957.

A COMPARATIVE FACTORIAL ANALYSIS OF SEMANTIC STRUCTURES IN MONOLINGUAL GREEK AND AMERICAN COLLEGE STUDENTS[1]

HARRY C. TRIANDIS
Cornell University
AND CHARLES E. OSGOOD
University of Illinois

OVER the past decade there has been growing interest in what is often called "the Weltanschauung Problem," the relation between language structure and cognitive processes (cf. 2, 9). To a very considerable extent, this interest has been stimulated by the writings of Whorf (cf. 14), who seriously questioned the commonly held notion that the cognitive processes of all human beings possess a common logical structure (which he called "natural logic") which operates prior to and independent of the particular language used to communicate. It was Whorf's thesis that linguistic patterns determine, to a large degree, how the individual perceives his world and how he thinks about it. He developed his hypothesis from the comparative analysis of a variety of languages, notably those of the Hopi, Shawnee, and Nootka Indians, against what he called Standard Average European (SAE).

The present paper is one of a series of studies supporting the opposite point of view—that there are certain aspects, at least, of cognition in human beings that are relatively independent of the particular language that they use to communicate. The aspect with which we are concerned is *meaning*, specifically the structure or organization of connotative meanings as revealed by factor analysis of the semantic judgments of concepts against descriptive scales. The first study of this nature was undertaken by Kumata and Schramm (4), comparing the semantic factor structures of bilingual Japanese and Korean students with those of monolingual American college students. The Japanese and Korean subjects (*S*s) were given translation-equivalent forms a few weeks apart, half of each group reacting first to the form in their native language and the other half to the English form; American monolinguals took the same English version, containing exactly the same concepts and scales, twice over the same interval. Not only were the factor structures for the bilinguals essentially the same regardless of the language used, and in fact as closely related as the test-retest structures of the American controls, but the semantic structures for both Japanese and Koreans corresponded very closely to those of the Americans. While this study demonstrates convincingly for these particular languages that the language structure per se does not influence the semantic organization of bilinguals, the fact that facile bilinguals were used limits conclusions about cross-cultural semantic similarities. In a subsequent study, however, Kumata (3) has been able to demonstrate a high degree of similarity in semantic structures between monolingual Japanese living in Japan and monolingual Americans. The present study extends this demonstration to monolingual Greek *S*s.

METHOD

In this investigation we have attempted to repeat in all essentials the procedures and materials used by Kumata in his monolingualism study. The ideal design for research of this type includes (*a*) the development of translation-equivalent forms of the semantic differential through the use of back translation by independent sets of bilinguals, (*b*) demonstration through the use of bilingual *S*s that language per se does not influence semantic structure, and (*c*) demonstration through the use of monolingual *S*s that language and cultural differences combined do not produce significant differences in semantic structure.

---

[1] The collection of the Greek data was made possible through a grant to William W. Lambert by the U. S. Public Health Service (No. 7162). The control group data collection was supported by a small grant to the first author by the Cornell Social Science Research Center. The analysis of these data was made possible through the facilities of the Institute of Communications Research and the Computer Laboratory, at the University of Illinois, under the direction of the second author.

In the present Greek study we omitted step (*b*), on the grounds that if languages as different as Korean or Japanese and English yielded no differences for the same bilingual *S*s, differences on this basis would not be expected for Greek vs. English. Furthermore, to the extent that similarities rather than differences in semantic structure prevail in step (*c*), differences in language structure per se are included in the demonstration. The method of back translation (step *a*) was used to insure translation equivalence between the English and Greek versions of the semantic instrument. Three bilinguals translated a large number of words from the English form into Greek and three *other* bilinguals translated the Greek terms back into English. Only those terms for which there was unanimous agreement were retained.[2]

Adaptation of Semantic Differential

The measurement technique used here was a form of the semantic differential as developed by Osgood and others (10, 12). The *S*s judge a set of concepts against a set of 7-step bipolar adjectival scales, such as *hard-soft*, *active-passive*, *pleasant-unpleasant*, *stable-changeable*, and so forth. To the extent that judgments along various scales are correlated, they may be analyzed for their factorial structure. *Evaluative*, *potency*, and *activity* factors have repeatedly been found by Osgood and his associates. This study, like Kumata's, began with a large number of scales previously used by Osgood and others and retained those that survived back translation. The final form consisted of 20 concepts and 30 scales which, when paired in all possible combinations, produced a 600-item test. Concepts were continuously rotated against scales rather than having *S*s make all of the judgments of a given concept on a single page. Except for the language, the forms used for Greeks and Americans were identical, with the same concepts, same scales, and same ordering; this procedure avoided the possibility of shifts in factorial structure due to concept-scale interaction (cf., 12, pp. 176). Table 1 gives the scales used in this study; the concepts used were as follows: FATHER, FEMALE, FAMILY, POLICE, MOUNTAIN, SICKNESS, LABOR UNION, WATERFALL, EARTHQUAKE, SOLDIER, JUSTICE, HOSPITAL, SNOW, DAWN, BIRD, FIRE, THIEF, BIRTH, SON, and OLD WOMAN.

Subjects

The "experimental" group consisted of 89 Greek monolinguals; the "control" group consisted of 43 American monolinguals. All *S*s were college students, the majority of the Greek students being recruited in the library of the student union of the University of Athens.[3] No Greek *S* had had as much as one full year of English; any who had learned English directly (e.g., from object to name-of-object in English rather than from Greek word to English word) were excluded from the sample. This control was necessary in view of Lambert's results (5) on the influence of the context of language learning on the semantic judgments of bilinguals. None of the American sample knew Greek. The Greek sample was 90% male as compared with only 60% males for the Americans; the age range for Greeks was 19–40 with a mode at 23, as compared with a range of 19–25 and a mode at 19 for Americans; more than a third of the Greeks were studying medicine and about half studying engineering, law, and mathematics, whereas the American *S*s were liberal arts students; it is also probable that the Americans were of somewhat higher socioeconomic status. We have no way of determining to what extent these differences may have contributed to the slight differences in semantic structure and the considerable differences in concept meanings to be reported.

Instructions

The instructions given to the American *S*s were identical with those described by Osgood and Suci (10). These instructions include a general statement of the purpose of the research (to determine what certain concepts mean to the *S*s), the meanings of the scale positions (Positions 1 and 7, *extremely*; Positions 2 and 6, *quite*; Positions 3 and 5, *slightly*; and Position 4, *equally*, *neither*, or *irrelevant*—all with respect to the polar terms) and the general attitude to be taken toward the task (working rapidly, giving first impressions, not looking back and forth in test, etc.). The instructions were modified slightly for the Greek *S*s—particularly (*a*) to indicate the cross-cultural nature of the study being undertaken and to enlist their support, and (*b*) to indicate willingness on the part of the experimenter to discuss the nature of the study after they had completed the task (about two-thirds of the *S*s took advantage of this opportunity).

RESULTS

Factorial Structures

The raw data for both Greek and American *S*s were punched into IBM cards, from which the means for each concept judged on each scale were determined. These means were transferred to punched tape, and all subsequent operations were done on the Illinois Digital Computer (ILLIAC). Interscale correlations for each group were computed from the concept means, i.e., cross-products taken

---

[2] This procedure resulted in the loss of 53% of the original scales and 13% of the original concepts. A good deal of this loss was due to the poor English of the bilinguals, who knew the meaning of only 60 or 70% of the words that were presented to them for translation.

[3] We wish to express our thanks to Lambros, Manager of the Student Union of the University of Athens, for making the arrangements for securing the *S*s.

TABLE 1

FIRST FIVE FACTORS FROM QUARTIMAX ANALYSIS

Greek Group					Scales	American Group				
1	2	3	4	5		1	2	3	4	5
.57	.53	−.30	.43	.04	High-low	.88	−.15	−.05	.08	−.25
−.02	−.80	−.33	−.07	−.16	Slow-fast	.13	.29	−.76	−.25	.04
.82	.34	−.03	.20	.08	Important-unimportant	.67	−.31	.14	−.42	.06
−.95	−.21	−.01	−.11	.02	Negative-positive	−.98	−.02	−.06	.10	.06
.93	−.07	−.07	.01	−.10	Grateful-ungrateful	.86	.31	.26	−.08	.10
−.65	−.03	−.28	.32	−.19	Heterogeneous-homogeneous	−.41	−.38	.21	−.55	−.01
.92	−.20	−.10	.16	.05	Good-bad	.94	.08	−.15	.06	.11
−.70	.36	.28	.18	−.05	Excitable-calm	−.39	.06	.72	.41	−.10
.88	.13	.24	.11	−.08	Optimistic-pessimistic	.90	.21	.08	−.06	−.05
−.97	−.12	−.03	.01	−.04	Unsuccessful-successful	−.96	−.01	−.04	.14	−.05
.91	.17	.13	−.17	.00	Timely-untimely	.91	−.02	−.07	.03	.22
−.91	.09	.07	.23	−.03	Egotistic-altruistic	−.78	.09	.33	−.22	−.11
.86	−.03	.15	−.17	−.19	Sociable-unsociable	.80	.24	.15	−.03	.16
−.84	−.25	−.02	−.22	−.22	Sickly-healthy	−.80	.03	−.30	−.13	.32
−.35	.70	−.35	−.00	−.35	Hard-soft	−.30	−.90	−.14	−.16	−.07
−.07	−.53	.62	−.03	.38	Light-heavy	.20	.83	.30	.24	.02
.77	−.27	−.10	−.23	−.21	Sober-drunk	.63	−.06	−.65	−.01	−.20
−.15	−.84	.74	.09	.08	Feminine-masculine	.07	.85	−.13	−.20	−.02
−.10	.69	−.10	−.40	−.31	Severe-lenient	−.42	−.75	.06	−.23	−.07
.37	.86	−.10	.01	−.09	Strong-weak	.49	−.76	.21	−.01	−.23
−.43	−.17	−.68	−.17	−.20	Old-new	−.25	−.25	−.70	.06	−.33
−.02	−.16	.17	.69	−.13	Colorful-colorless	.51	.19	.50	.05	−.61
−.61	.09	−.12	.33	.42	Unusual-usual	−.52	−.04	.33	−.17	−.53
−.19	−.36	.44	.12	−.43	Private-public	.13	.64	−.20	−.45	−.05
.55	.73	.17	−.10	.09	Active-passive	.25	−.32	.78	−.05	.13
−.19	−.06	−.64	−.05	.52	Cold-hot	.09	−.06	−.47	.65	−.01
−.72	−.24	.46	.07	.27	Changeable-stable	−.51	.32	.69	−.12	.02
.09	.33	−.06	−.73	−.00	Angular-rounded	−.41	−.83	−.02	.10	.19
−.77	.05	−.06	−.52	−.28	Ugly-beautiful	−.83	−.29	−.09	−.19	.35
.52	.61	.40	−.33	.01	Straight-curved	.28	−.80	.04	.18	.38

across the set of 20 concepts for each pair of scales.[4]

The correlational matrices were first factored by Thurstone's centroid method (13). The first factor was clearly evaluative for both groups, including such scales as *high-low*, *important-unimportant*, *positive-negative*, *good-bad*, *optimistic-pessimistic*, *timely-untimely*, and the like, and accounted for 43% of the total variance in the Greek group and 39% in the American group. The second factor, accounting for 17% and 16% of the variance respectively, appeared as a "dynamism" factor (confluence of potency and activity dimensions), e.g., *fast-slow*, *excitable-calm*, *hard-soft*, *heavy-light*, *masculine-feminine*, *severe-lenient*, *strong-weak*, and *active-passive*. Factors 3 to 5 were less readily interpretable, accounting for 8%, 8%, 5% and 17%, 5%, 5% in the two groups, respectively, and the remaining factors all accounted for less than 4% of the variance.

The first five factors accounted for at least 80% of the total variance in both groups. These factors were rotated by the Neuhaus and Wrigley (7) quartimax method, and indices of factorial similarity were computed (Wrigley and Neuhause, 16) The results are given in Tables 1 and 2. Accepting an index of about .75 as demonstrating significant agreement, it can be seen that the first two factors are related well beyond this level. But whereas the Americans display the typical evaluation-potency-activity pattern as their first three factors, this is not as true for the Greek *S*s. Something like a "receptivity" factor (cf., Osgood, Suci, and Tannenbaum, 12, pp. 64) appears as Factor 4 for the Greeks and Factor 5 for the Americans (e.g., *colorful*, *unusual*, *rounded*, *beautiful*, *curved*), and some factor

---

[4] To save printing costs, the table of intercorrelations has been deposited with the American Documentation Institute. Order Document No. 5636 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for 35 mm microfilm or $1.25 for 6 by 8 photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 2

INDICES OF FACTORIAL SIMILARITY FROM QUARTIMAX ANALYSIS

Greek Group 1	2	3	4	5	American Group 1	2	3	4	5		
1.00	.23	.08	−.04	−.01	.93	−.05	−.12	.16	.20	1	Greek
	1.00	−.08	−.13	−.12	.09	−.81	.39	−.03	−.03	2	
		1.00	.05	.09	.08	.40	.60	−.15	.21	3	
			1.00	.20	.23	.38	.33	−.19	−.70	4	
				1.00	.03	.25	.27	.54	−.14	5	
					1.00	.15	−.10	.14	.01	1	Americ
						1.00	−.02	.01	−.09	2	
							1.00	−.22	−.22	3	
								1.00	.04	4	
									1.00	5	

TABLE 3

FIRST THREE FACTORS FROM SQUARE ROOT ANALYSIS

Greek Group			Scales	American Group		
Timely-Untimely	Light-Heavy	Slow-Fast		Timely-Untimely	Light-Heavy	Slow-Fast
.46	−.48	−.20	High-low	.73	−.08	−.14
−.20	.06	.98	Slow-fast	.17	−.05	.98
.73	−.32	−.08	Important-unimportant	.49	−.31	−.11
−.87	.18	−.02	Negative-positive	−.84	−.03	.03
.77	−.04	.23	Grateful-ungrateful	.79	.18	.27
−.64	−.18	.01	Heterogeneous-homogeneous	−.33	−.33	−.15
.74	.06	.29	Good-bad	.94	.06	.10
−.63	−.09	−.50	Excitable-calm	−.43	.15	−.47
.91	.10	−.05	Optimistic-pessimistic	.86	.23	.03
−.92	.08	−.06	Unsuccessful-successful	−.88	−.03	.01
1.00	−.00	−.00	Timely-untimely	1.00	−.00	.00
−.80	.02	−.25	Egotistic-altruistic	−.71	.05	−.15
.77	−.08	.19	Sociable-unsociable	.68	.21	−.05
−.82	.00	.14	Sickly-healthy	−.68	−.05	.24
−.20	−.70	−.40	Hard-soft	−.24	−.88	−.17
−.06	1.00	.00	Light-heavy	.19	.98	.00
.69	.10	.40	Sober-drunk	.70	−.26	.32
−.28	.42	.55	Feminine-masculine	.07	.69	.44
.31	−.53	−.41	Severe-lenient	−.41	−.66	−.16
.47	−.63	−.46	Strong-weak	.37	−.57	−.37
−.48	−.41	.26	Old-new	−.24	−.42	.29
−.15	.21	−.12	Colorful-colorless	.31	.39	−.33
−.69	.02	−.26	Unusual-usual	−.64	.04	−.13
−.26	.23	.23	Private-public	.04	.27	.40
.66	−.29	−.54	Active-passive	.19	−.07	−.80
−.22	−.14	.21	Cold-hot	.19	−.12	.25
−.66	.43	−.03	Changeable-stable	−.54	.49	−.26
.19	−.17	−.07	Angular-rounded	−.34	−.69	−.32
−.62	−.16	−.01	Ugly-beautiful	−.70	−.38	.01
.66	−.04	−.41	Straight-curved	.28	−.68	−.36

difficult to name, but somewhat corresponding (e.g., *cold* and *public*), appears as 5 for the Greeks and 4 for the Americans.

In an attempt to see if even further agreement in factorial structure could be demonstrated, three common pivotal variables were selected from the original correlational matrices as being highly loaded on independent factors but minimally correlated with each other, and a square-root factor analysis (cf. Wrigley, 15) was applied to the same correlations. The scales thus selected as pivots wer *timely-untimely*, *light-heavy*, and *slow-fast*. Th logic of this procedure is that if nearly as muc of the total variance can be accounted for b these arbitrarily chosen factors as by the firs three factors of the centroid, then this i approximately as good an orthogonal rotatio

TABLE 4

INDICES OF FACTORIAL SIMILARITY FROM SQUARE ROOT ANALYSIS

Greek Group 1	Greek Group 2	Greek Group 3	American Group 1	American Group 2	American Group 3	
−1.00	−.20	−.03	.89	−.12	−.12	1
	1.00	.40	−.02	.83	.29	2 Greek
		1.00	.22	.34	.91	3
			1.00	.13	.12	1
				1.00	.22	2 American
					1.00	3

The proportions of total variance accounted for by the square-root analysis were 61% (Greeks) and 61% (Americans) for the three arbitrary factors; proportions of variances accounted for by the first three factors of the centroid were 68% and 73% respectively. This arbitrary rotation into evaluative, potency, and activity factors is thus not quite as successful for either group as the original analysis. Table 3 gives the factor loadings for the square-root analysis. On the other hand, as can be seen from inspection of the indices of factorial similarity in Table 4, the correspondences across Greek and American language/culture groups are significantly high for all three factors. The clearest indicators of the evaluative factor in both groups are *timely-untimely*, *positive-negative*, *optimistic-pessimistic*, *successful-unsuccessful*; the clearest indicators of the potency factor are *heavy-light*, *hard-soft*, *severe-lenient*, and *strong-weak*; the clearest indicators of the activity factor are *fast-slow*, *excitable-calm*, and *active-passive*.

Having demonstrated considerable correspondence in general semantic factorial structure across Greek and American groups, we may now ask if there are any particular scales of judgment that display clear differences in usage. We may take the three square-root factors as reference vectors and see if certain scales run through this space in decidedly

TABLE 5

MEAN SCORES OF CONCEPTS ON EVALUATIVE AND POTENCY FACTORS AND ON SELECTED SCALES, FOR GREEK AND AMERICAN GROUPS

	Father		Son		Family		Old Woman		Female		Birth		Hospital		Sickness		Justice		Soldier		
	G[a]	A[b]	G[a]	A[b]	G[a]	A[b]	G[a]	A[b]	G[a]	A[b]	G[a]	A[b]	G[a]	A[b]	G[a]	A[b]	G[a]	A[b]	G[a]	A[b]	
E-factor +	2.3	2.5	2.6	2.7	2.3	2.1	4.3	4.4	3.1	2.8	2.5	2.0	2.3	2.4	4.7	5.4	2.2	2.4	2.6	3.6	E-factor −
P-factor +	2.9	2.8	3.5	3.3	3.5	3.5	5.0	4.9	5.0	5.3	3.6	3.4	4.0	3.1	3.6	4.0	2.7	2.2	2.7	2.2	P-factor −
Sociable	2.0	2.0	2.2	2.6	1.6	1.6	3.4	4.0	1.9	2.1	2.3	3.1	1.3	3.7	3.8	4.8	1.8	3.1	2.5	3.2	unsociable
Slow	5.3	4.6	5.5	5.0	4.5	4.1	2.1	1.8	4.4	4.3	4.1	2.8	4.8	4.1	4.2	3.2	4.9	3.3	6.1	5.4	fast
Excitable	4.5	4.5	3.8	3.1	5.1	4.0	4.4	4.3	3.7	2.6	4.4	2.0	5.8	4.7	3.7	3.5	5.4	5.0	4.1	3.6	calm
Feminine	6.6	6.7	6.4	6.5	3.6	4.0	2.3	1.9	1.4	1.2	3.4	2.1	3.9	4.1	3.4	3.5	4.1	4.8	6.5	6.7	masculine
Colorful	3.9	2.7	3.8	2.7	3.8	2.3	4.2	4.8	3.4	1.9	4.0	2.4	3.8	4.8	4.1	5.3	3.9	4.0	4.1	2.7	colorless
Unusual	5.4	5.4	5.5	4.7	5.7	5.7	5.5	5.4	5.2	4.4	6.0	4.9	5.7	5.1	5.6	4.4	5.9	5.2	5.8	5.5	usual
Old	3.7	3.3	6.6	5.3	3.9	2.9	1.5	1.3	5.9	4.9	4.0	5.0	4.9	4.3	3.3	3.0	3.8	2.4	5.4	4.6	new
Private	2.7	2.7	2.9	3.1	2.5	1.7	3.3	2.6	3.3	2.7	3.1	2.0	5.8	4.8	3.7	2.7	6.4	5.9	6.3	5.8	public
Cold	4.9	4.3	4.8	4.3	5.3	4.9	3.4	2.9	5.7	4.9	5.2	5.3	5.1	3.3	4.5	4.7	4.6	3.0	4.6	3.7	hot
	Police		Labor Union		Thief		Snow		Dawn		Bird		Fire		Waterfall		Earthquake		Mountain		
	G[a]	A[b]	G[a]	A[b]	G[a]	A[b]	G[a]	A[b]	G[a]	A[b]	G[a]	A[b]	G[a]	A[b]	G[a]	A[b]	G[a]	A[b]	G[a]	A[b]	
E-factor +	2.4	3.3	2.4	3.5	4.0	5.0	3.2	3.1	2.8	2.3	3.2	3.0	3.7	3.6	3.4	2.8	4.4	4.8	3.4	3.2	E-factor −
P-factor +	2.7	2.2	3.3	2.6	3.4	3.4	4.8	5.5	4.7	5.1	5.2	5.6	3.3	3.9	2.8	3.4	2.7	1.8	2.4	1.8	P-factor −
Sociable	2.0	3.9	1.7	3.4	5.3	5.8	3.9	2.8	3.3	3.6	3.5	3.5	3.7	3.1	3.7	3.3	4.3	5.3	4.0	4.3	unsociable
Slow	5.9	5.2	4.9	4.1	6.1	5.8	3.6	3.3	3.8	2.8	6.3	5.9	4.9	5.8	6.0	5.8	5.7	5.5	3.8	2.4	fast
Excitable	4.4	4.1	4.0	2.2	3.6	2.5	4.7	4.7	5.2	4.7	4.5	3.1	2.8	2.0	2.9	2.6	3.0	1.5	4.9	5.6	calm
Feminine	5.1	6.6	3.7	5.7	4.9	5.8	3.7	2.8	3.2	2.5	3.5	2.5	3.3	4.3	4.9	3.8	4.9	5.4	4.3	5.8	masculine
Colorful	4.1	3.3	3.9	3.9	4.0	3.8	4.7	3.0	2.7	1.6	2.3	1.9	2.0	1.6	3.9	2.4	4.1	2.9	2.9	2.3	colorless
Unusual	5.9	5.9	5.7	5.0	5.2	3.6	5.1	4.9	5.6	4.4	5.4	5.3	5.5	4.3	4.7	3.9	5.0	2.3	5.4	4.5	usual
Old	3.7	3.3	3.9	4.3	3.4	3.5	4.4	5.2	4.8	4.8	4.1	4.9	3.8	3.6	3.0	2.7	3.4	3.5	2.6	1.7	new
Private	6.2	6.2	5.5	5.5	4.0	4.0	4.7	4.7	4.4	3.6	4.6	4.9	4.2	4.3	5.2	4.6	4.7	5.9	5.7	4.7	public
Cold	4.2	3.3	4.8	4.3	4.1	3.5	1.5	1.4	3.6	3.5	5.2	4.4	6.7	6.8	2.2	2.0	4.2	5.0	2.6	2.1	hot

[a] Greek group.

[b] American group.

different directions. The scale *high-low*, for example, is almost purely evaluative for Americans but relates equally to evaluation and potency for Greeks. On the other hand, the scale *heterogeneous-homogeneous* is largely evaluative for the Greeks (with *homogeneous* being "good") while it is relatively independent of these factors for the Americans. Although *severe-lenient* is clearly a potency scale for both groups, for Americans *severe* tends to be "bad" and for Greeks *severe* tends to be "good." *Active-passive* is a pure activity scale for Americans, but for Greek *S*s being *active* is clearly "good" as compared with being *passive*. Finally, we note that both *angular* and *straight* (as compared with their opposites, *rounded* and *curved*) tend to be evaluatively "good" for the Greek *S*s; whereas they are less so, or even "bad," for Americans; both of these scales are primarily potency variables for Americans, but not so for Greeks. The remaining 24 of the 30 scales show no marked differences between cultures.

Concept Meanings

Space does not permit tabulation of all of the 20 × 30 × 2 means of concepts on individual scales for both groups, but Table 5 presents the most important evidence. The six scales previously shown to be used differently by the two culture groups have been eliminated, since differences here could be attributed either to concept meanings or scale meanings. All mainly evaluative scales are represented by a single E-factor score (obtained by averaging the most purely evaluative scales according to the square-root analysis, *timely-untimely*, *positive-negative*, *optimistic-pessimistic*, and *successful-unsuccessful*); this may be thought of as an attitude score (cf., Osgood and Tannenbaum, 11). Similarly, a P-factor (potency scales, *heavy-light*, *hard-soft*, and *strong-weak*) can be represented in a single value. On the basis of our previous work (Osgood, Suci, and Tannenbaum, 12), in which differences as small as ⅓-scale unit in factor scores under test-retest conditions were found to be significant at the .05 level, we shall assume conservatively that differences in factor scores here of ½-scale unit are significant. In the remainder of Table 5 are presented a number of individual scales, roughly clustered, in which significant differences[5] for at least some concepts are obtained. Differences as large as approximately 1.0-scale units may be accepted as indicative of a real semantic difference.

DISCUSSION

Generality of Semantic Structure

The major hypothesis of this research was that there are dimensions of connotative meaning that are independent of differences in language and culture. Its confirmation requires that, with the concepts judged and the scales used in judgment held constant and translation equivalent, the same or highly similar dimensions of meaningful discrimination be demonstrated across two language/culture monolingual groups. In the present study, we have compared the semantic structures obtained from Greeks with those obtained from Americans when judging 20 varied concepts against 30 varied scales. An unrotated Thurstone centroid factor analysis of the data indicated both an "evaluative" factor and a "dynamism" factor for both groups, these factors accounting for 60% and 55% of the total variance in the Greek and American groups, respectively, and showing significant correspondence in content across groups. An arbitrary rotation in terms of scales representing three factors regularly found in our earlier work yielded evidence for "evaluation," "potency," and "activity" factors in both groups, these factors accounting for 61% of the total variance and again being significantly related across groups. We interpret this as evidence for a high degree of semantic generality across Greek and American language/culture communities.

When we add these results to those already obtained by Kumata (3) for monolingual Japanese vs. Americans, the case for complete generality of semantic structure despite differences in language and culture is considerably strengthened. Greek belongs to a subgroup of the Indo-European family of languages that is quite distinct from all other members of the family. Our present findings, then, support the assumption that the same general semantic structure will be found in all Indo-European languages. Japanese (and Korean, for which we have data on bilinguals, at least) is a non-

---

[5] All differences on single scales as large as one scale unit are significant at least at the .05 level.

Indo-European language, and the similarity of Kumata's results to those obtained here supports an extension of our hypothesis of semantic structural similarity to languages in general. Data have been collected on several Indian groups of the American Southwest (cf., Casagrande, 1), including Hopi, Zuni, and Navajo, and if similar correspondences can be demonstrated there, the case for complete generality of basic dimensions of meaningful discriminations across the human species will be strong indeed.

The Problem of Cross-cultural Comparisons

Comparisons across cultures, despite differences in language, are obviously quite feasible for material traits, like the making of pottery, or even some nonmaterial traits, like family organization. But comparisons in terms of many nonmaterial traits—particularly those relating to attitudes, briefs, values, and the like—are extremely difficult when different languages are used. How can we assess and compare those things for which language is the chief expression if there are no common yardsticks? Indeed, if the Whorfian thesis were taken as absolute, comparisons of this sort would be ruled out—and this is exactly the position ("cultural relativity") taken by some cultural anthropologists and linguists. If, on the other hand, it can be shown that there is some common ground in meaning systems across language communities, if certain common dimensions, a common semantic frame of reference, can be demonstrated for human beings in general, then comparison becomes possible because we have some identifiable reference points. One of the long-term goals of our work in semantic measurement is to demonstrate the feasibility of such cross-language, cross-culture comparisons, with the purpose of improving cross-cultural communication and understanding.

Comparison of Greek and American Nonmaterial Cultures

Although this study was not designed as an anthropological investigation, and our bases for comparison are therefore not extensive, we do have a small amount of incidental evidence on the differences between Greek and American nonmaterial cultures that may illustrate the use of cross-cultural semantic measurement. Having demonstrated the existence of certain common factors, and certain commonly used scales within this factor system, we can use these as a reference frame within which to interpret the differences in either scale meanings or concept meanings that appear.[6]

Let us look first at *scale meanings*. In the first place, it is worth noting that the Greeks give somewhat more weight to the evaluative factor than do the Americans. This finding held for the unrotated centroid analysis, the quartimax rotation, and the arbitrary rotation into three generally dominant factors. The Greeks, in other words, tended to use all scales of judgment somewhat more evaluatively. By way of comparison, the same tendency was even more marked in the Korean bilinguals (but not the Japanese bilinguals) studied by Kumata and Schramm (4). Such relatively greater use of the evaluative dimension probably represents a tendency toward what Korzybski refers to as dichotomous or "Aristotelian" thinking. The Greeks, like the Koreans, seem to reflect a more "militant" frame of reference, as will be seen, and it is interesting to speculate on whether rotation toward a "black-and-white" judgmental system generally results from such cultural pressure.

The differences in meanings of certain individual scales seem to reflect such a "militant" frame of reference. *High* is potent as well as good for Greeks, but mainly good for Americans. *Severe*, *active*, *homogeneous*, *angular*, and *straight* are all favorably evaluated as compared with their opposites by the Greeks, but not for the Americans—in fact, *lenient*, *rounded*, and *curved* tend to be the more favorable poles in the latter case. One could, of course, make separate interpretations of these scale differences—for example, that Americans have to "put up with" a much greater flux of heterogeneity than Greeks and hence have developed a more tolerant, less evaluative usage. And with respect to the positive evaluation of *straight*, we find Lee (6) saying:

---

[6] The first author, being reasonably familiar with both Greek and American cultures, has ventured to make interpretations of some of these differences. However, he is not trained as a cultural anthropologist and his interpretations must be judged accordingly.

> The beloved is commonly likened to a tree . . . most commonly to a cypress—slim, straight, tall, hard, resistant—"I embrace a cypress tightly" sings a lover. . . . Firmness and straightness are extremely desirable personal qualities. Babies were swaddled to keep them from growing crooked and to make their backs firm. The gesture of mother love is one of holding a firm, stiff, straight bundle against the breast, not the crooking of the arms to accommodate a cuddling body. Firm and straight is also the ideal personality. . . . When Eleutherios Venizelos (one of the leading statesmen of Greece) died, the leading Athenian newspaper wrote: "The cypress is fallen."

Reflecting the fact that *activity* has a more positive evaluative tone for Greeks, we find Lee saying:

> Folk tales make fun of the lazy wife who comes to grief. To call a girl diligent, is to say something about her attractiveness. A poem which school children used to memorize spoke of the joyous little housewife, a little girl who got up early to set the house in order. There was nothing of the repellent goody-goody about her, and nothing strange about calling industriousness joyous.

The close correspondence between such anecdotal or observational evidence of cultural anthropologists and the results obtained with the semantic differential would seem to strengthen the claims of both to validity.

Differences in *concept meanings* are striking in some cases and in others practically nonexistent. We may use differences as large as ½ a scale unit for E- and P-factor scores and approximately 1.0-scale unit for individual scales as conservative criteria of significance (see Table 5). Looking first at certain family concepts, we see that there are no differences whatsoever in the meaning of OLD WOMAN; the concepts FATHER and SON are viewed as more *colorful* by American *S*s, but SON as more *new* to the Greeks; the FAMILY is considered more *calm* by the Greeks, but more *colorful* and *private* by the Americans. In part, these differences may represent a general tendency on the part of Americans to "romanticize" everyday things and events, as compared with the Greeks who seem to take a more realistic view. This is particularly clear for FEMALE, who is seen as much more *excitable* and *colorful* by Americans, as well as somewhat more *unusual* and *warm* (one merely needs to open any "slick" American magazine to note the repetitious reinforcement of this conception); however, FEMALE is judged *older* by Americans than Greeks. We shall note this tendency of Americans to romanticize again in connection with natural objects.

The cluster of concepts relating to vital life processes (BIRTH, HOSPITAL, and SICKNESS) reveal some very interesting differences between cultures. Americans give a more favorable evaluation to BIRTH and see it as *faster*, more *excitable*, and more *colorful*; but, whereas they view it as a *private* affair, definitely associated with *feminine* roles, Greeks view it as a more *public*, *sociable*, family event, *old* and familiar, *not* at all *unusual*. Whereas in America we have the customs of baby presents to the wife, baby showers, handing out of cigars and the like, in Greece the relatives merely convene, look at the baby, and congratulate the parents. But if the Greek attitude toward BIRTH is matter-of-fact, not so their attitude toward HOSPITAL. They view the HOSPITAL as definitely more *sociable*, *colorful*, *warm*, and *public* than do Americans. The presence of large numbers of relatives and the somewhat less rigid rules of the Greek hospital probably contribute to this picture. The Americans, on the other hand, see the HOSPITAL as a more forbidding, coldly efficient place—for them it is more *potent* (*hard*, *heavy*, and *strong*), *cold* and threatening (*excitable*), and they view it as an intimate *private* affair rather than a public familiar one. It is interesting that although the Americans view SICKNESS less favorably (E-factor), they also see it as less potent or dangerous (P-factor)—these differences possibly reflecting the fact that whereas Americans consider more minor ailments as "sickness" and are more confident in medical ability to handle them, the Greeks associate the term with major illnesses. And here again we find the now familiar pattern in which the Greeks view SICKNESS as more *sociable*, *public*, *colorful*, and ordinary (*usual*). The extended family usually lives geographically close together and all the relatives flock to the bedside of the patient.

Another general difference between Greek and American "world views" seems to be that whereas the former see social institutions and roles in a more personal, concrete way, the latter see them as impersonal and abstract. This contrast appears particularly for the concepts JUSTICE, SOLDIER, POLICE, LABOR UNION, and to lesser extent THIEF. Americans judge all of these concepts *colder* and *less sociable* than do the Greeks, and in most cases significantly so. Americans also have a decidedly less favorable attitude (E-factor) toward these things, with the single exception of JUSTICE; yet, with the exception of THIEF, Americans view these things as more powerful (P-factor). In other words, whereas Americans tend to see the law, the military, unions, and those who represent them as remote, but powerful and rather threatening, the Greeks see them in terms of the individual humans who are playing the parts. The soldier is an individual, not a role, as is the policeman and the thief. The Greek who has trouble with the law is more likely to look for some powerful relative or friend to support him than for a good lawyer. As Lee says, "The traditional Greek pattern . . . is that individuals work not collectively but within an intricate web of interdependence and mutual aid. The abstract end, unless it is implemented in terms of persons, is meaningless" (6, p. 112). The environment of the American factory, with its cold, impersonal atmosphere, has attracted very few Greek immigrants, while the restaurant business has felt a major impact due to immigration from this country. It is also worth noting that LABOR UNION, POLICE, and THIEF are judged significantly more *feminine* by Greeks; whether this simply reflects facts (that women play these roles more frequently in Greece) or represents lesser connotative potency is difficult to say. The much more favorable attitude of the Greek *S*s toward LABOR UNION probably is due to both the more favorable press Greek unions receive and the fact that the University of Athens students are of somewhat lower socioeconomic status than the Cornell students.

We have already noted the tendency of American *S*s to romanticize common everyday natural events. This trend is particularly true of physical objects and events. SNOW is more *sociable*, *feminine*, and *colorful*, as well as less potent (i.e., *light*, *soft*, and *weak*) for Americans than for Greeks; DAWN is also thought of as more favorable and less potent, as well as being more *feminine*, *colorful*, *unusual*, and *slow*; BIRD is more *excitable*, *warm*, *feminine*, and *new*; WATERFALL is likewise both more liked (E-factor) but less potent (P-factor), while being more *feminine*, *colorful*, and *unusual* for Americans. For the concept FIRE we find an abrupt shift from this pattern; although FIRE is judged more *potent* and *usual* by the Greeks, it is also judged significantly more *feminine*—a result that suggests some specific metaphorical connotation.

Finally, we have two aspects of the physical environment, EARTHQUAKE and MOUNTAIN, where we might reasonably expect somewhat different experiences (i.e., referents) for our two populations. EARTHQUAKES, even though of a generally mild nature, are rather common events in Greece; they are very rare experiences for most Americans (particularly in the East) and acquire their meaning more through legend (e.g., the San Francisco quake) than through direct involvement. Accordingly, we find Americans judging them as more *potent*, *excitable*, *unusual*, *colorful*, *hot*, *public*, and *unsociable* than the Greeks. Similarly, Greece is ruggedly mountainous throughout, but they are not towering mountains nor do they form great contrasts with extended plains; most Cornell students have either seen or heard about the towering Rockies and Sierras, in contrast with the more gentle hills and plains with which they are familiar. In any case, we find American *S*s judging MOUNTAIN as more *potent* but *slow* (massive and majestic?), more *unusual* and more *masculine* and *old*; Americans also judge MOUNTAIN as being more *public*, perhaps because of the association with national parks, sight-seeing, and the like.

SUMMARY

This is one of a series of studies on the cross-cultural, cross-language generality of connotative semantic factors. The judgments of 20 concepts against 30 descriptive, bipolar scales (translation equivalent) by a sample of monolingual Greek college students were compared with the judgments of the same materials by a sample of American college students. Factor analysis of these data indicated a high degree of similarity in the basic semantic dimensions used by both groups. With this common reference frame, it was possible to show how Greeks and Americans differ in their usage of certain individual descriptive scales and in their meanings for certain concepts. For the Whorfian hypothesis, the present results imply that certain aspects of human cognition are relatively independent of the structure of the language used to communicate. The semantic differential would appear to be adaptable for

the needs of cultural anthropology in the cross-cultural study of similarities and differences in nonmaterial culture.

REFERENCES

1. CASAGRANDE, J. B. The Southwest Project in comparative psycholinguistics: A progress report. Soc. Sci. Res. Council, *Items*, 1956, **10**, 41–45.
2. HOIJER, H. (Ed.) Language in culture. *Amer. Anthrop.*, 1954, **56**, 6.
3. KUMATA, H. A factor analytic investigation of the generality of semantic structure across two selected cultures. Unpublished doctoral dissertation, Univer. of Illinois, 1957.
4. KUMATA, H., & SCHRAMM, W. A pilot study of cross-cultural methodology. *Pub. Opin. Quart.*, 1956, **20**, 229–237.
5. LAMBERT, W. E. Studies of the verbal behavior of bilinguals. Paper presented at the Midwestern Psychological Association, May, 1956.
6. LEE, DOROTHY D. Greece. In Margaret Mead (Ed.), *Culture patterns and technical change*. UNESCO, 1953.
7. NEUHAUS, J. O., & WRIGLEY, C. F. The quartimax method: An approach to orthogonal simple structure. *Brit. J. Stat. Psychol.*, 1954, **7**, 81–92.
8. OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 197–237.
9. OSGOOD, C. E., & SEBEOK, T. A. (Eds.) *Psycholinguistics: A survey of theory and research problems*. Suppl., *Internat. J. Amer. Ling.*, 1954, **20**, Memoir 10.
10. OSGOOD, C. E., & SUCI, G. J. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50**, 325–338.
11. OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychol. Rev.*, 1955, **62**, 42–55.
12. OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univer. of Illinois Press, 1957.
13. THURSTONE, L. L. *Multiple-factor analysis*. Chicago: Univer. of Chicago Press, 1947.
14. WHORF, B. L. *Language, thought and reality: Selected writings of Benjamin Lee Whorf*. (Edited by Carroll, J. B.) New York: Wiley, 1956.
15. WRIGLEY, C. F. A rapid method of factor analysis for a large correlation matrix: The modified square-root procedure. *Amer. Psychol.*, 1956, **11**, 415.
16. WRIGLEY, C. F., & NEUHAUS, J. O. The matching of two sets of factors. *Amer. Psychol.*, 1955, **10**, 419.

Received July 26, 1957.

CONDITIONING OF ATTENTION AS A SOURCE OF AUTISTIC EFFECTS IN PERCEPTION[1]

RICHARD H. WALTERS

University of Toronto

THERE have been numerous experiments designed to demonstrate the influence of motives, needs, and values on perceptual processes. One of the main criticisms of the "directive-state theory of perception," to use F. H. Allport's term (1), has been that the experiments which supposedly support such a theory have been frequently carried out under conditions that facilitate misperception or perceptual distortion. In other words, techniques such as tachistoscopic exposures, reduced illumination, and other devices that weaken stimulus conditions have been habitually employed. In the present experiment such techniques were avoided, thus bringing the stimulus conditions more in line with those encountered in everyday life.

Murphy (2, 3) has been long concerned with autistic effects in perception, i.e., the "movement of cognitive processes in the direction of need satisfaction." Demonstrations of autistic effects by Murphy and his associates have been previously carried out for the most part by providing monetary rewards or presenting electric shocks or other pleasant and noxious stimuli. Such extrinsic rewards and punishments should not, however, be essential for the production of autistic effects. In the present experiment, an attempt was made to demonstrate perceptual learning of an autistic kind by the use of a problem-solving device as a training task; no rewards and punishments were provided other than those associated with the subject's (*S*'s) successes and failures during the course of his efforts to solve the problem. The exact nature of *S*'s motivations under such conditions may not be specifiable; presumably, his habitual modes of responding in a problem-solving situation will be tapped.

The experiment was first carried out in 1954, using two groups of *S*s, one composed of 14 business college students and psychiatric aides at Topeka, Kansas, and another composed of 19 undergraduates from an introductory psychology course at Stanford University, making 33 *S*s in all. The experiment was repeated, using improved controls, in the summer of 1956; on this occasion 32 Stanford undergraduates were used as *S*s.

METHOD

In the 1954 experiment, the test objects were 12 complex multicolored designs, similar to those used in the Gottschaldt test. Each was accompanied by a simple outline drawing of a figure which was identical in outline with part of the corresponding complex design. The designs differed, however, from those used in the Gottschaldt test in that the simple figure could be located in either of two areas. In half the designs, one of the areas containing the simple figure was colored red; the other was colored green. In the other half of the designs, the simple figure could be located either in a blue or in a yellow area.

Procedure

The experiment commenced with a pretraining test in which *S* was merely asked to find the simple figure in the more complex one. His "solution" was recorded in terms of the color of the area in which he perceived the simple figure.

The training period was designed to reward one color of each pair used on the test cards and to punish the other. Since there were two pairings of colors, for each *S* two colors were rewarded during the training period and two were punished. The rewarded colors were blue and green; the punished colors were red and yellow.

The training task was selected as unlikely to have any apparent connection with the task used in the test periods, and it was presented as another problem to be solved. The training material consisted of six decks of cards, 12 cards in each deck; each card was red, green, yellow, blue, orange, or violet. The latter two colors were neutral colors from the point of view of the experiment, i.e., they were neither rewarding nor punishing. The cards were numbered on the backs as follows: in the first deck, 1-1, 1-2, 1-3, up to 1-12; in the second deck, 2-1, 2-2, 2-3, up to 2-12; the remaining decks were correspondingly numbered, up to 6-12. On the colored sides of the cards were instructions to *S*.

---

[1] The research was part of the Perceptual Learning Project under the direction of Gardner Murphy, and was supported in part by the National Institute of Health, United States Public Health Service, Project M-715, and by the Menninger Foundation. The author is grateful to Gardner Murphy for his continued stimulation and encouragement during the planning and execution of this experiment, and to Robert C. Pierce for his assistance during its replication.

When *S* called the number of a card, he was shown the colored side with the instruction.

The *S* was presented with a checker board on which a path was marked leading from the top left-hand square concentrically around the board to the center. He was instructed as follows:

> Look at this checker board with the path marked out on it. I am going to start you at this corner of the board and you have to go to the center—so [*E* demonstrates]. Your moves on the board will be determined in this way. Here I have some decks of cards. Each of the cards has an instruction on it, saying whether you are to move forward, go backward, or just stay put. These cards are of different colors. If you can call the right colored cards you can move forward all the time.
>
> These cards are numbered. [The numbering system was explained at this point.] Now these cards are arranged according to a system, and your task is to discover the system. You are to call the number of a card, and I will then raise the card (so), and show you the instruction, which you will then follow. When you have solved the problem, you will be able to call the right colors all the time and so move forward and reach the center of the board. Each time you call a card I shall remove it from the pack and you will not be able to use it again until we start a new trial. You will have to call the right colors all the time in order to get to the center.
>
> You may be wondering what I mean by a system in this case. Well, a very simple system might be that every fifth card in a deck enabled you to move forward, so that you would have to call 1-5, 1-10, 2-3, and so on. This is a more complex system, however, and remember, there are always two numbers involved in designating any card.
>
> When you have called a certain number of cards, we shall put them all back in their decks again, so that you can begin once more, but you will get only a limited number of trials during the testing period. We are going to make a note of the time you are working on each trial.
>
> At first you won't have any hunches what to do and you will have to proceed by trial and error. But after a few calls you may get a hunch and you can go on trying that one out to see if it works. Work as fast as you can, for unless you do you won't be able to solve the problem in the time we give you.

The "system" was set up so that "rewarding" cards were numbered by a pair of numbers, the smaller of which was a factor of the larger, e.g., 2-4, 2-6, 3-6, 3-9, 6-1. By excluding from the "system" Cards 1-7 through 1-12, on the grounds (if *S* challenged) that the largest multiplier which could be used was 6, the number of decks available, the "system" allowed for 25 rewarding cards on which were written the instruction, "Move forward two" or "Move forward three," so that calling all the 25 cards would allow the *S* to move through sufficient squares to reach the center. Another 25 cards had the instruction, "Move backward two" or "Move backward three," while the remaining 22 cards instructed *S* to "Stay put."

Each time *S* called the number of a card, *E* held up the card at a table-width distance from *S*, so that the instruction had to be read in a way which would allow the color of the card to have some impact on *S*. Three or four trials were given each *S*, the total training period taking 40 to 50 minutes, according to the speed with which *S* made his calls. In order to keep motivation reasonably high, assistance was given *S* if he showed no sign at all of making progress.

Immediately following the training period, *E* informed *S* that he would like to have him try out the first task again to see if he could improve his time on a second trial. The reference to "improving the time" was included primarily to supply a reason for the retest, not in order to hurry *S*. The *S*'s choice was again recorded in terms of the color of the area in which he located the simple form.

The *S*s were then questioned by *E* concerning their opinions on the nature and purpose of the experiment. The purpose of this questioning was to determine whether or not *S*s became aware of the function of the colors of the cards during the experiment.

When the experiment was replicated in 1956, some of the more obvious defects in the procedure were remedied. Sixteen test cards were now used, and care was taken to distribute the *S*s so that half of them were rewarded by the presentation of green and blue cards; half of them by the presentation of red and yellow cards. This modification necessitated the use of two sets of training cards which were identical except for the location of the colors within the decks. In addition, an attempt was made to eliminate the possible influence of the relative positions of the colored areas on the test cards by presenting one half of the *S*s with one of the longer edges of the test cards at the top, and the other half with the other of the longer edges at the top, i.e., with the card rotated through 180°. Thus, eight *S*s were run under each of four possible conditions.

Since the objection could be raised to the earlier experiment that any change in response resulting from training might be due merely to the relative frequency with which the rewarded cards were presented, *E* now introduced into the testing situation an assistant who recorded every presentation during the testing period in terms of the color of the card. This made it possible to assess the relationship between extent of change and the frequency of presentation of the rewarded cards.

Finally, the questioning procedure which followed the experiment was standardized by the presentation of a mimeographed booklet of questions to which *S*s gave written answers. The questions were as follows:

1a. What do you think we were concerned about in this experiment?

1b. What were we trying to find out?

2. What did you think the first part of the experiment was designed to test (i.e., the one with the simple and complex figures)?

3. What do you think the other part was designed to test (i.e., the one with the cards and checkerboard)?

4. Suppose you were told you had just been through a single experiment, that the two parts were not separate experiments, and that we were just trying to find out one thing, what would you think we were trying to find out?

5a. Did you notice how the colors of the cards in the twelve packs were related to your moves on the checkerboard?

5b. If so, how?

6. Now, what do you think we were trying to find out?

The questions were arranged in the booklet so that, by not allowing S to turn a page before answering the questions on it, he was forced to answer earlier questions before seeing later ones. Thus, there was an opportunity to discover how complete or superficial his understanding of the procedure might be, and how much this understanding might have influenced his manner of responding during the experiment.

RESULTS

It is possible to present the results of these experiments in two ways: (*a*) A comparison of the number of *S*s whose net change in response (over the 12 or 16 figures) was in a rewarded direction with the number of *S*s whose change was predominantly in a punished direction; and (*b*) in terms of the significance of the difference between the mean number of responses to rewarded (or punished) colors in the pretraining period and the mean number of responses to rewarded (or punished) colors in the posttraining period.

Tables 1 and 2 give the findings by both these methods; the benefit of a one-tailed test has been claimed in each case. It can be seen that the 1954 result was significant only when a nonparametric test was used. Failure to obtain significance using a t test was due to the fact that changes in the rewarded direction were for the most part slight, whereas a few *S*s showed large changes in the punished direction. The extent of changes among the 1956 group was generally quite small, and there were no changes in the punished direction greater than 2.

For the 1956 group, the excess of presentations of rewarded over punished cards was calculated for each S and this score was correlated with the extent to which the *S*s changed in the rewarded direction. The correlation (Pearson's r) was $-.38$, which for an N of 32 is significant at the .05 level. This relationship is in the opposite direction from that which would be expected if frequency of presentation were the major factor in producing change. The correlation between the extent of change and the total number of presentations of training cards (a possible index of extent of training) was negligible ($r = .04$). The interpretation of these results is difficult because we do not know how the neutral cards functioned for individual *S*s; for some they may have been as punishing as a card which required them to retrace their steps on the board, but for others they may have been merely nonrewarding items. Some such factor may in part account for the rather surprising significant negative correlation reported above. It seems quite evident, nevertheless, that the effect that was produced was not simply a function of the excess of presentations of rewarded cards, nor was it simply a function of the total length of the training period.

TABLE 1

NUMBER OF SUBJECTS CHANGING IN A REWARDED OR IN A PUNISHED DIRECTION FROM PRETRAINING TO POSTTRAINING TRIALS

	Rewarded Direction	Punished Direction	No Change
A. 1954 Group			
	21	8	4
chi square =	5.82	$P < .01$	
B. 1956 Group			
	16	5	11
chi square =	5.76	$P < .01$	

TABLE 2

COMPARISON OF MEAN NUMBER OF RESPONSES TO REWARDED CARDS IN PRETRAINING AND IN POSTTRAINING TRIALS

	Pretraining Mean	Posttraining Mean	*SD* diff.	t	P
1954 Group $N = 33$	5.67	6.09	2.36	1.03	$>.05$
1956 Group $N = 32$	7.81	8.78	1.98	2.77	$<.01$

An analysis of the protocols obtained from *S*s in the 1956 experiment indicated that not one of them was aware of the purpose of the experiment. Even after being asked Question 4, most *S*s did not succeed in meaningfully connecting the two parts of the experiment. The majority of *S*s still stated that the experiment was concerned with learning, memory, or ability testing. The nearest approach to indicating the true purpose of the experiment was found in six *S*s who thought at this stage the E must be testing for interference or transfer effects, but even these *S*s did not realize the reward-punishment function of the cards. Only 11 *S*s indicated that they noticed the colors of the cards, and only six of these were able to identify correctly the colors of the cards that allowed them to move forward or backward (Question 5). Question 6 failed to elicit any new ideas from the majority of *S*s; of those who added something, five thought that the experiment dealt with the ability to observe, four now thought it was designed to see if color aided in learning, and one other thought it dealt with color relationships (not further specified).

DISCUSSION

Positive results on both experiments leave little doubt that the technique produces autistic effects. The technique appears to have certain advantages over those generally used in experiments having a similar purpose. Not only are the conditions for perceiving brought closer to those encountered in everyday life, but in addition the use of a problem-solving situation as a training device successfully obscures the purpose of the experiment from *S*.

By contrast we may consider some other techniques that have been used, e.g., when *S* is presented with one object in the training situation, he is given money; when he is presented with another object, money is taken away. If he wants to please *E*, he has been given plenty of hints on how to respond when placed in an ambiguous situation. In the present situation, however, *S* is primarily concerned with proving himself. He wants to get winning cards because, if he does, he is demonstrating his competence. This is the set with which he enters the posttraining trial also. The protocols clearly indicate that at no time was *S* consciously looking for the color which *E* wished him to choose. Given the continuity of motivation throughout the experiment and *S*'s lack of awareness of the purpose of the experiment, we can now explain the autistic effect in terms of a conditioning process.

Let us suppose that initially *S* has a certain set of response tendencies when placed in the pretraining situation. These may be position tendencies, e.g., he may be more likely to look at the top right-hand corner of the card than at any other part of it. Or there may be tendencies to look at certain colored areas rather than at others, tendencies related perhaps to color preferences, but nevertheless ones built up out of past experiences in somewhat similar situations. The *S* is then trained by a reward-punishment procedure to call certain numbers. Whenever *S* calls a "correct" number, he is presented by *E* with a card instructing him to proceed toward the center of the board. The experimental situation is, in fact, a social situation, and, following Sears (4), we may regard *E*'s act as an environmental event that enables *S* to make the goal-response of moving forward along the board. *S*'s expectation of this environmental event may be regarded as an anticipatory goal response which is itself strengthened during the training period. Associated with *E*'s presentation of a card is *S*'s perception of a certain color, and this response on the part of *S* may be thought of as also moving forward in time in the sequence of responses in the form of a fractional anticipatory goal response. The training procedure thus, we assume, increases not only the probability of *S*'s calling "correct" numbers but, also, the probability of his making certain anticipatory motor and receptor adjustments. There may also be anticipatory neural adjustments produced by our training procedure which would serve as stimuli for the perception of the rewarded colors in situations where these are not associated with the number-calling procedure. The results of questioning *S*s after the experiment indicate, however, that such anticipatory responses, if they occur, must be largely subliminal in character.

The view presented here is that the autistic effects produced in this experiment result from the conditioning of attention. Attention may be thought of predominantly in terms of motor adjustments, or in terms of a complex of motor and neural adjustments of an anticipatory nature, which would serve as stimuli for the perception of goal objects. In any case, the effects produced by the training procedure can be quite simply explained by the principle of generalization. The total stimulus complex that is present throughout the experiment remains relatively constant. This complex includes the subject-experimenter relationship (possibly defined by *S* as an examiner-examinee relationship), the kind of instructions that *E* provides, and the materials employed by *E*, i.e., colored cards. Moreover, *E*'s instructions are such that they may be expected to arouse anticipatory responses relating to the perception of color both in the training and in the test trials. It is presumably these responses that are conditioned during the training period and serve as stimuli resulting in a change in overt response in the posttraining period.[2]

---

[2] Robert C. Pierce and Jan R. Pierce, in a subsequent unpublished study based on the experiments reported

The directive-state theory of perception has been developed largely to account for the occurrence of nonveridical and idiosyncratic percepts. The present experiment is not of a kind that induces such effects, but in this respect it is no different from some other experiments in which ambiguous or reversible figures are employed. The theory does, however, claim that the occurrence of nonveridical elements in perception results from the operation of rewards and punishments, and regards experiments similar to the present one as supporting the theory by showing how the operation of rewards and punishments may result in *S*'s perceiving one figure rather than another in situations where both might be perceived. In view of what we know about the operation of rewards and punishment on responses other than perceptual ones, it would be indeed strange if they produced no effects on perceptual behavior. What, then, is the characteristic which gives a response an autistic character? Murphy (2) has already provided a suggestion that has been largely overlooked in discussion of experiments designed to demonstrate autistic effects. He writes: "Typical autism . . . is an expression of those phases of third level perception in which the interaction of affective and cognitive elements goes on without the observer's being sharply aware of what is happening." In this sense, the effects produced in the present experiment are clearly autistic in character. The modifications of *S*'s anticipatory responses—to return to the terms in which the results of this experiment have been discussed—are subliminal in character.

This interpretation of autistic effects raises many problems which are out of the scope of the present paper. Since much of our learning seems to occur at a subliminal level (as an example consider the process of identification of the child with significant adults in its environment), the inevitable conclusion would seem to be that the effects of much of our learning, perceptual or otherwise, must be autistic in character. This conclusion, however, loses its point when we also regard even pathologically autistic phenomena as behavior patterns which are subject to the general laws of learning but occur only under conditions that allow for learning at a subliminal level and under strong reinforcement.

SUMMARY

The purpose of this study was to demonstrate the possibility of producing autistic effects in perception in a situation that rather closely paralleled everyday conditions in which objects are perceived. The experiment was first carried out in 1954 and was replicated with slight improvements in 1956.

Subjects were presented with the task of locating a simple figure in a more complex colored figure in a test similar to the Gottschaldt test. The simple figure could be located in one of two areas within the complex figure; these areas were of different colors. A training period followed in which, by means of a problem-solving device, some colors promoted success (reward) and others failure (punishment). Finally, the Gottschaldt-like figures were re-presented, and the changes in *S*'s responses were recorded.

Both the 1954 and 1956 experiments produced positive results. The *S*s showed a significant change, resulting in an increase of responses to rewarded colors in the posttraining trial. It was shown in the 1956 experiment that the change was not simply a function of relative frequency of exposure to cards of different colors.

It was clear from *S*s' replies to a questionnaire concerning their ideas of the nature and purpose of the experiment that they were not aware of the association of the colors with reward and punishment processes. An attempt was made to explain the results in terms of subliminal anticipatory responses. It was suggested that the autistic quality of the change produced resides primarily in *S*'s lack of awareness of the effects of the training or conditioning process.

REFERENCES

1. ALLPORT, F. H. *Theories of perception and the concept of structure.* New York: Wiley, 1955.
2. MURPHY, G. *Personality: A biosocial approach to origins and structure.* New York: Harper, 1947.
3. MURPHY, G. Affect and perceptual learning. *Psychol. Rev.*, 1956, **63**, 1–15.
4. SEARS, R. R. Social behavior and personality development. In T. Parsons & E. A. Shils (Eds.), *Toward a general theory of action.* Cambridge: Harvard Univer. Press, 1952. Pp. 465–478.

Received July 17, 1957.

---

here, used two sets of Gottschaldt-type designs. The two sets of cards corresponded in the color but not in the shape of the designs. The one set was presented prior to training, the other set in posttraining trials. Under these circumstances, *S*s failed to condition. This finding may necessitate some revision of the theoretical analysis offered here.

SELF-IDEAL CONGRUENCE AS AN INDEX OF ADJUSTMENT[1]

RALPH H. TURNER AND RICHARD H. VANDERLIPPE
Oberlin College

UNDERLYING the use of the *Q*-sort technique is the theoretical assumption that the satisfaction or concern of an individual with his phenomenal self is an operationally definable datum of great importance in the area of adjustment. The discrepancy between the placements of a given characteristic on the self scale and the ideal scale has been taken as indicating not only the way in which an individual perceives himself as possessing this characteristic but the degree to which he values this trait. Given these assumptions, then successful therapy should be accompanied by a reduction in the magnitude of self-ideal discrepancies. Further, self concepts should change more as a result of therapy than ideal concepts since the latter are firmly anchored as general societal concepts, whereas the former may be more idiosyncratic. These expectations have been confirmed by Butler and Haigh (3).

Nevertheless, the confirmation of these expectations is not itself a sufficient reason for assuming the adequacy of the basic theoretical framework within which the hypotheses were developed. Dymond (4) has pointed out that the therapist's attitude of satisfaction with a case may be communicated to the individual, who then reflects these attitudes in his self-description. Even if this possible source of contamination did not exist, it would still be necessary to test the assumption that a change of the phenomenal self in the direction of the ideal self may be taken as an indication of better adjustment. Because this is difficult to demonstrate on a before-and-after basis with the same *S*s, an alternative is to select two groups who presently represent the before-and-after conditions as defined by self-referent statements and to obtain behavioral criteria of adjustment for each. The need for such an approach has become even more pressing since Taylor (7) has shown that successive *Q* sorts, without benefit of therapy, reflect changes of the kind reported for successful counseling. Brownfain's study (2) was a step in this direction, but the relationship between his instrument and the conventional *Q* sort requires exploration.

METHOD

Measurement of Self and Ideal Self

One hundred and seventy-five upper-class students enrolled in introductory psychology at Oberlin College served as subjects (*S*s) on a volunteer basis. The 100 items and general methodology of the SIO *Q* sort developed by Butler and Haigh (3) were selected because of their frequent and systematic use in current research. Each of the items of self-reference was placed on a 3 x 5 card to facilitate manipulation. The *S*s were seen in groups of 20, each *S* having a set of 100 items, an instruction sheet, and sufficient table space for privacy and easy manipulation of the items. The following instructions were read to the *S*s:

> You are about to take part in a preliminary test in connection with a research project in the department of psychology. It is very important that you follow directions as carefully as possible. You have before you a pile of 100 cards. Each card has a single sentence written upon it. After I have finished reading these instructions, you are to take the cards and sort them into nine piles in such a way as to describe yourself as you see yourself today. The nine piles will range from those statements that are least like you (Pile 1) to those statements that are most like you (Pile 9). There is a specific number of cards to be placed in each pile. This number is shown on page 2 of these instructions. You can interpret the meaning of each statement in any way you think is correct. I will answer no questions concerning meaning during the sorting. It is very important that you sort these cards as honestly as you can. It is not necessary to attempt arranging the cards within each pile. After you have finished sorting the cards, check to make sure you have the correct number of cards in each pile.

When this self sort was concluded, the items were shuffled and directions given for the ideal sort. These were similar to the instructions given previously except for the substitution of the phrase "sort the cards to describe your ideal self—the kind of person you would most like to be."

Following the mode of analysis adopted by Brownfain (2) and by Hanlon, Hofstaetter, and O'Connor (6), the *S*s were divided into two groups, the high comprising the 25 *S*s having the highest correlations between self and ideal sorts and the low comprising the 25 having the lowest correlations. This division permitted an approximation to the differences that have been reported for groups seeking therapy as opposed to groups not seeking therapy. No evidence was found

---

[1] The writers wish to acknowledge the contributions of Elizabeth Walker and Bruce Norton to several phases of this research. L. E. Cole supervised a portion of this study and served as a consultant.

that men and women college students differ significantly in the extent of their self-ideal correspondence. This finding is in agreement with that of Block and Thomas (1), who combined the data for their *S*s. Similar practice was followed here, yielding a high group of 11 men and 14 women and a low group of 10 men and 15 women. It is in terms of the significance of differences between each of these two groups of 25 *S*s on various criteria of adjustment that the validity of the *Q* sort and the theoretical assumptions upon which it rests were tested. For this purpose, Fisher's *t* for small samples was utilized. All values are based upon one-tailed tests of significance since a directional hypothesis governed the original design of the experiment.

The Q-*Adjustment Score Criterion*

Dymond's (4) procedure was followed in utilizing the 37 items of the SIO *Q* sort that trained clinical psychologists believed a well-adjusted person would say were unlike him (Piles 1–4) and the 37 items they believed a well-adjusted person would say were like him (Piles 6–9). The *Q*-adjustment score for any *S* was the number of these 74 items that were sorted into the "correct" side of the distribution in the course of making a *Q* sort of the self.

Behavioral Criteria of Adjustment

The behavioral criteria against which the *Q*-sort scores were tested were health, extracurricular participation, sociometric indices, and scholastic adjustment. From the files of the student health service, three indices of health were computed: (*a*) mean number of days per college year spent in the hospital, (*b*) mean number of times per college year that assistance for any reason was requested, (*c*) the total number of negative health items mentioned in the pre-entrance health history as ascertained by a form on which both the student and the family physician had entered appropriate data.

The extracurricular participation score for all *S*s in the high and low groups was calculated by first compiling a list of 94 extracurricular organizations. Through interviews with officials of each organization, every position within the organizations was given three ratings on five-point scales for the amount of time involved, the status of the position within the particular group, and the status of the position on the campus as a whole. The sum of these three ratings constituted the weight given that position. The number of years a position was held was multiplied by its appropriate weight, and all such products for each activity summed to yield a total score for each *S*. This was divided by the number of years *S* had been on the campus to obtain the final participation score.

The sociometric indices involved getting from dormitory housemothers the names of the nine students whose rooms were closest to each *S* in the high and low groups. Each of the ten students in these living units received a ranking form with the names of the other nine written as column headings to facilitate marking. By entering numbers in the appropriate name column, each student ranked the other nine on each of the following questions: (*a*) With which person do you spend the most time when you are in a mood to relax? (*b*) Which person would be best at organizing a group of about 12 people for a particular purpose (such as a clean-up campaign)? (*c*) If you wanted to talk about a personal problem with a trusted and helpful person, with whom would you be most apt to talk? (*d*) Which person is the quietest? (*e*) Which person do you feel has had the most favorable influence upon you? (*f*) Which person makes friends with others most quickly? (*g*) Who would most fairly evaluate the circumstances under which an important college regulation had been broken and decide what punishment, if any, should be given? (*h*) Who is the most cooperative teamworker, working well with a group and doing his/her share? (*i*) Who is most apt to handle wisely any situation which calls for quick, clear thinking? (*j*) Which person is most apt to initiate spontaneous group recreational activities? (*k*) With whom do you (or would you) especially want to be in a dormitory living group again? The mean rank each *S* received from other members of his living group was then computed for each of these items.

The comparative scholastic achievement of the two groups was obtained by comparing the mean cumulative academic grades of those in the high and low groups. Because such a comparison might reflect a difference in initial scholastic potential rather than a difference in adjustment, the mean of the raw scores on the American Council Test (8) was computed for each group. This test, taken by all students during the first week of the freshman year, was found to yield a product-moment correlation of .51 with cumulative grades for the students on this campus. It thus fulfills the function of providing a moderately valid index of scholastic aptitude. To the extent that the two groups do not differ significantly with respect to scholastic aptitude, any significant difference in scholastic achievement may be interpreted as an indication of a difference in scholastic adjustment.

Standard Tests as Criteria

Although not an original part of the study design, 50 of the original *S*s were given the Guilford-Zimmerman Temperament Survey (5) as part of another project. Under standard administration and scoring procedures, scores were obtained for the 10 traits measured. Product-moment correlations were run between raw scores for each of these traits, on the one hand, and the correlations between self and ideal *Q* sorts, on the other, the latter being first transformed into corresponding z scores.

RESULTS

Table 1 presents the range of the self-ideal correlations in order to facilitate a comparison between the high and the low groups. By use of the z transformation, the mean correlations for each group were calculated. The correlation of .79 indicates reasonably high correspondence between self and ideal ratings for the high group. The corresponding r of .11 for the low

TABLE 1

CORRELATION BETWEEN SELF SORT AND IDEAL SELF SORT

Group	N	Range	Mean z	Equivalent r
All *S*s	175	−.08 to .91	.63	.56
High	25	.75 to .91	1.08	.79
Low	25	−.08 to .25	.11	.11

TABLE 2

MEAN GROUP DIFFERENCES IN HEALTH INDICES

Health Index	High Group	Low Group	Diff.	t Value	$p<$
Days in hospital	1.14	.93	.21	.458	.35
Reports to clinic	3.99	2.71	1.28	.977	.20
Health history items	1.24	1.84	.60	.470	.35

group does not permit rejection of the hypothesis that the population correlation is zero.

The *Q*-adjustment scores for the high group yielded a range of 44–59 with a mean of 52.04. The low group yielded a range of 13–43 with a mean of 29.68. This latter mean corresponds closely to that of 28.80 reported by Dymond (4) for a group that had sought therapy. The mean for the high group is somewhat above the 44.96 she reported for her control group. The range of adjustment represented by the present high and low groups, as measured by this criterion, is certainly no smaller than that found between those seeking therapy and those not seeking therapy. Since the distribution of *Q*-adjustment scores for the high and low groups did not overlap, there is a significant relationship between self-ideal *Q*-sort scores and this external criterion of adjustment.

Table 2 indicates that no evidence was found that the high and low groups differed significantly with respect to the number of days spent in the hospital, the number of times the student sought assistance at the health service, or the number of health items mentioned in the pre-entrance health history.

The mean social participation score for the high group was 20.24, while that for the low group was 13.84. A test of the significance of the difference yielded a t of 2.12, significant beyond the .05 level. It appears that those with substantial correlations between self and ideal ratings on the *Q* sort are significantly more active as measured by this index of social participation than are those whose correlations do not differ significantly from zero.

TABLE 3

MEAN RANKS IN SOCIOMETRIC ITEMS[a]

Item	High Group (N − 24)	Low Group (N − 18)	Diff.	t Value	$p<$
All	4.2	5.2	1.0	2.38	.03
a	4.6	5.1	0.5	1.34	.10
b	3.7	5.1	1.4	2.50	.01
c	4.4	5.1	0.7	1.75	.05
d	4.9	4.6	0.3	.51	.35
e	4.2	5.3	1.1	2.46	.01
f	4.4	5.3	0.9	1.67	.10
g	4.0	5.0	1.0	1.82	.05
h	4.0	4.7	0.7	1.39	.10
i	4.1	5.3	1.2	2.17	.03
j	4.1	5.8	1.7	3.06	.01
k	4.5	5.3	0.8	1.80	.05

[a] A low rank is considered an indication of better adjustment except in the case of item *d*.

Table 3 summarizes the data from the sociometric study. The original two groups of 25 each suffered some attrition because of lack of cooperation on the part of some members of the living group. The mean rank received by *S*s from those in the living group is indicated for each of the questions and for all items combined. With the combined rank taken as a score for the instrument as a whole, those in the high group achieve a significantly better rating than those in the low group. Examination of the individual items indicates that those in the high group are given preferential ranks without a single exception. Eight of the 11 individual items yielded an acceptable level of significance.

The cumulative academic grade, based upon a point system currently in effect, yielded a mean of 1.78 points for the high group and 1.52 for the low group. This difference, which favors the high group, was found to reach the criterion of significance ($p < .05$) with a t of 1.81. The mean raw scores of the high and low groups on the American Council Test were 134.6 and 133.1 respectively. This difference was not significant ($p < .40$ with a t of .31). The scholastic superiority of the high group thus does not appear to rest primarily upon an initial difference in scholastic potential and suggests somewhat better scholastic adjustment on the part of those whose self-ideal congruence is high.

Table 4 indicates the correlation between self-ideal congruence expressed as z values and the traits measured by the Guilford-Zimmerman Temperament Survey (5). Five of these

TABLE 4

CORRELATIONS BETWEEN SELF-IDEAL CONGRUENCE EXPRESSED AS z VALUES AND TRAITS MEASURED BY THE GUILFORD-ZIMMERMAN TEMPERAMENT SURVEY

Trait	Correlation with Self-Ideal Congruence
1. General Activity	.50*
2. Restraint	−.10
3. Ascendance	.58*
4. Sociability	.36*
5. Emotional Stability	.36*
6. Objectivity	−.03
7. Friendliness	−.16
8. Thoughtfulness	.41*
9. Personal Relations	−.25
10. Masculinity	.10

* Significant at the .01 level.

yielded correlations sufficiently large that the hypothesis of zero correlation could be rejected at the .01 level. Those high in self-ideal congruence tended to have high scores in general activity (strong drive, much energy, enthusiasm), ascendance (strong self-defense, leadership, ease in speaking and persuading), sociability (at ease with others, readily established rapport), emotional stability (optimism and freedom from neurotic tendencies), and thoughtfulness (mental poise, observation of behavior in others, capability as supervisor of others, reflectiveness). In each instance, the scores for those high in self-ideal congruence are indicative of better adjustment than are those for *S*s low in self-ideal correspondence.

DISCUSSION

Of the criteria of adjustment utilized, a number of significant differences were found between the group high in self-ideal congruence and the group low in this respect. In no instance was a significant difference found in a direction contradictory to the hypothesis. The failure of certain criteria to show a significant difference between the two groups may reflect the inadequacy of these criteria or may indicate that self-ideal congruence is not equally valid as an index of adjustment when applied to all dimensions of adjustment. The possibility of an approach in terms of dimensions of adjustment is suggested by noting that the *Q*-adjustment score was found to be the external criterion of adjustment most closely related to self-ideal congruence. Since this instrument was the only one in the enviable position of having all its items drawn from the same statements used in the *Q* sort, it preserved a close relationship with it in terms of the dimensions tapped.

Health-related behavior can occasionally be shown to reflect personal conflict. It may be questioned that it is an appropriate index of adjustment within the adjustment range of the student population sampled. The overlay of genuine organic complaints constitute determiners of behavior that may effectively mask adjustment-motivated behavior in this area.

Block and Thomas (1), contending that self-satisfaction is not linearly related to the social dimension of adjustment, quite properly suggested that it may, as a consequence, have only very limited relevance as an index of the outcome of psychotherapy. It nevertheless appears that in the context of adjustment defined by the criteria utilized here, the theoretical assumptions upon which the conventional *Q* sort are based received substantial support.

SUMMARY

A test of the theoretical assumptions and the methods upon which the conventional *Q* sort rests was made using several criteria of adjustment. The results tended to support the theoretical assumptions underlying the *Q* sort as currently utilized. Although certain criticisms of the procedures involved are not thereby completely invalidated, it would appear that they are not sufficiently serious to justify abandoning either the method or the assumptions supporting the method.

The emergent composite picture of the college student high in self-ideal congruence (as contrasted with the student low in self-ideal congruence) is that of one who participates more in extracurricular activities, has a higher scholastic average, is given higher sociometric ratings by his fellow students, and receives higher adjustment ratings on both the *Q*-adjustment score and certain traits measured by the Guilford-Zimmerman Temperament Survey.

REFERENCES

1. BLOCK, J., & THOMAS, H. Is satisfaction with self a measure of adjustment? *J. abnorm. soc. Psychol.*, 1955, **51**, 254–259.
2. BROWNFAIN, J. Stability of the self-concept as a dimension of personality. *J. abnorm. soc. Psychol.*, 1952, **47**, 597–606.

3. BUTLER, J. M., & HAIGH, G. V. Changes in the relation between self concepts and ideal concepts consequent upon client centered counseling. In C. R. Rogers & Rosalind Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. 55–75.
4. DYMOND, ROSALIND. Adjustment changes over therapy from self-sorts. In C. R. Rogers & Rosalind Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. 76–84.
5. GUILFORD, J. P., & ZIMMERMAN, W. S. *The Guilford-Zimmerman Temperament Survey: Manual of instructions and interpretations.* Beverly Hills, Calif.: Sheridan Supply Co., 1949.
6. HANLON, T. E., HOFSTAETTER, P. R., & O'CONNOR, J. P. Congruence of self and ideal self in relation to personality adjustment. *J. consult. Psychol.*, 1954, **18**, 215–218.
7. TAYLOR, D. M. Changes in the self concept without psychotherapy. *J. consult. Psychol.*, 1955, **19**, 205–209.
8. THURSTONE, L. L., & THURSTONE, THELMA G. *Psychological examination for college freshmen: 1947 norms.* Washington, D. C.: American Council on Education, 1948.

Received July 19, 1957.

SOME EFFECTS OF FEEDBACK FROM THE EXPERIMENTER ON CONFORMITY BEHAVIOR[1, 2]

EDWARD E. JONES, H. HERBERT WELLS, AND RICHARD TORREY

Duke University

EXPERIMENTAL investigations of conformity behavior have typically dealt with contrived situations of cognitive conflict in which the subject (*S*) is forced to choose between sensory information and discrepant information provided by consensus. A host of different variables have been shown to condition this choice, among them being the nature of the stimulus (4), the unanimity and size of the majority (1), whether or not the judgments are made anonymously (5, 6), *S*s' set or orientation to the task (5, 9), and attitudinal or personality variables that *S* imports into the situation (6). One variable hitherto relatively neglected is the provision of information from a third source which is authoritative and presumably quite independent of consensus.

Crutchfield (4) has reported briefly the results of an attempt on the part of *E* to reinforce conformity. The *S* is presented with a perceptual comparison task, sees that other *S*s have agreed on an (incorrect) answer, and then makes his own judgment. After *S* makes his judgment, *E* reports as correct the alternative chosen by the rigged majority, regardless of *S*'s choice. According to Crutchfield,

> The effect of this so-called "correction" method is striking. As the series of judgments goes on, these individuals express greater and greater conformity to the group pressure on slides which are of the same character as those for which earlier in the series the false group consensus was thus reinforced by the false announcement by the experimenter (p. 198).

In line with Schein's findings in an experiment on human imitation (7), Crutchfield found little tendency for this "learned imitation" to generalize to problems of a different type (e.g., attitudes and opinions).

While these findings are eminently plausible, it is not clear whether this is merely a simple learning situation in which continuous reinforcement is important, or whether the information provided by *E* leads to abrupt cognitive restructuring. The latter would be the case, for example, if the *S* was rather quickly led to reassess the situation and then cast his lot with consensus from that point on. We might better speak of reinforcement, however, if the tendency to yield or not to yield were related to the number of reinforcements of conformity or independence. The first experiment was conducted to shed some light on this question.

EXPERIMENT I

If *E*'s report is used by *S* to make a quick assessment of the situation, then *S* should be just as responsive to a few reports from *E* as to a great many. For this reason, the frequency and direction of reinforcement were independently varied in this first experiment. Thus, in a situation where groups of *S*s were exposed to identical stimulus figures and on a number of critical trials were incorrectly informed of the judgments of others before announcing their own, *E* either sided with consensus after the judgments were made or confirmed the answer that was actually correct. This feedback was either provided on every trial or only up to and including the first trial on which the consensus gave an incorrect response.

The design called for five experimental groups: (*a*) *Continuous feedback, pro-consensus*—*E* always sides with consensus, even when consensus is incorrect. (*b*) *Discontinued feedback, pro-consensus*—*E* sides with consensus up to and including the first trial where there is discrepancy between consensus and "reality." At this point *E* says, "you're on your own" and provides no further feedback. (*c*) *No feedback.* (*d*) *Discontinued feedback, pro-"reality"*—identical to (*b*) except that *E* gives the correct answer the first time there is discrepancy between consensus and accuracy and then provides no further feedback. (*e*) *Continuous feedback, pro-"reality"*—*E* always reports

---

[1] This study was conducted in the program of the Organization Research Group of the Institute for Research in Social Science at the University of North Carolina. The O.R.G. is supported jointly by the Ford Foundation and the Office of Naval Research [Nonr-855 (04)]. The research was partially financed by a grant to the senior author from the Duke University Research Council.

[2] We are very much indebted to Milton Rosenbaum for building the apparatus and constructing the stimulus materials.

the correct answer, regardless of what the consensus is.

While it is obvious that more conformity would be expected when *E* sides with group consensus than when he sides with reality, two further questions motivated the first experiment: (*a*) In comparison to the no-feedback control group, which condition has the most (positive or negative) effect on conformity, feedback favoring consensus or feedback favoring "reality"? (*b*) Does continuous feedback have a greater effect on conformity or independence than feedback which is discontinued after the first critical trial?

Method

Apparatus and stimulus materials. The apparatus used in both experiments to be reported here was similar in design to that reported by Crutchfield (4) and Deutsch and Gerard (5). Each group was composed of four female *S*s who were separated by partitions that prevented them from seeing each other. Before each *S* was a panel containing four rows and three columns of lights. At the base of the panel there were three noiseless switches, each of which would light up a given light in the same row. Each *S* was ostensibly assigned a different row (A, B, C, or D) and was instructed to indicate her judgment, when her turn came, by pressing the button corresponding to her choice (1, 2, or 3). This would illuminate the appropriate light on her own panel and, presumably, others' panels. Actually, each *S* was privately informed that she was Subject D and it was therefore fairly plausible when *S* found that her turn to make a judgment always came last. As in experiments employing similar apparatus, information concerning the judgments of A, B, and C were presented to each of the actual *S*s by an experimental assistant operating a master panel in an adjacent room. Thus on every trial, each *S* was exposed to a standardized pattern of prior judgments presumably being made by the other three *S*s.

The judgments to be made in all cases involved the observation of airplane silhouette slides presented at brief exposures. Each slide consisted of a standard silhouette and three comparison silhouettes, one of which was identical to the standard. *S*'s task was to locate the identical comparison version and to identify it by pressing the appropriate button at the base of her panel. Twenty different slides were shown to the *S*s at 2-sec. exposure. The same series was repeated without comment or delay at 1-sec. exposure for a total of 40 trials per session. On Trials 4, 7, 10, 12, 15, 17, 19, 20, 24, 27, 30, 32, 35, 37, 39, and 40, the bogus judgments relayed into the panels by the experimental assistant were unanimously in error. Since the slides were originally placed in a random order, we assume that neutral and critical trials were equally difficult for the *S*s.

Subjects. Female members of an introductory psychology class served as *S*s for the experiment. While the design called for three groups of four *S*s in each of the five experimental conditions, one *S* could not be used because of apparatus failure, leaving a total of 59 *S*s whose data are reported.

Orientation to the experiment. Each group of *S*s was told that the ultimate purpose of the experiment concerned the development of efficient methods for training and evaluating aircraft observers and that there was considerable interest in the use of females in the Ground Observer Corps. The apparatus was described to the *S*s, and they were told, "Because we don't want to turn this into any kind of a competition between friends, we have arranged it so that each person will know her letter [row] designation but not anyone else's. For this reason, the machines are not letter-designated in the order they are placed next to each other." Each *S* was then given a slip of paper bearing the letter designation "D," and throughout the experiment, D was always instructed to respond last.

After the last judgment trial, a postexperimental questionnaire was administered followed by revelation of the conditions and actual purpose of the experiment. While occasional *S*s claimed to have been suspicious during the later trials, this had little if any discernible effect on their actual yielding behavior. All *S*s were very cooperative in pledging and maintaining secrecy.

TABLE 1

Mean Frequency of Yielding as a Function of Feedback Conditions

Feedback Condition	Mean	*SD*
I. Continuous pro-consensus	9.67	6.00
II. Discontinued pro-consensus	7.64	4.11
III. No feedback	7.50	5.42
IV. Discontinued pro-"reality"	3.42	2.32
V. Continuous pro-"reality"	1.96	1.62

Results and Discussion

Disregarding the effects of feedback for a moment, the amount of conformity in general was rather striking. Fifty-four out of 59 *S*s yielded at least once, and there was yielding on 37 per cent of all critical trials. When this latter figure is compared with the 6 per cent of errors made on otherwise equivalent non-critical trials, it is clear that there are some general effects of social consensus in the present situation.

Our main interest lies in the effects of various kinds of feedback from *E*. Table 1 shows the means and standard deviations for frequency of yielding by condition. There are rather clear-cut differences between conditions, reflected in a decline in both mean and standard deviation as one goes from pro-consensus feedback to pro-"reality" feedback. Because of the significant heterogeneity of variance, analysis of variance was inappropriate, and the data were analyzed in terms of individual pairs of comparisons, using a t test presuming unequal population values of σ (10).

The present data do not give strong support to Crutchfield's contention that reinforcement of consensus by *E* results in striking increases in conformity. Condition I (continuous pro-consensus feedback) does not result in significantly more yielding than Condition III (no feedback). However, Condition V (continuous pro-"reality" feedback) differs from III at the .01 level. Essentially the same results were obtained using nonparametric median χ^2 tests. Under the conditions of this experiment, then, it seems fair to conclude that feedback from *E* which reinforces independence is much more effective in changing the level of conformity than feedback which reinforces conformity.

Comparisons of the two conditions of early but discontinued feedback with their respective paired conditions of continuous feedback show differences in the amount of yielding which are not significant in either case. Condition II results in a frequency of yielding almost identical to Condition III, the control. Condition IV, on the other hand, results in significantly less yielding ($P < .05$) than Condition III and almost significantly more yielding than Condition V ($P < .10$).

The present data are equivocal with regard to a decision between a continuity-reinforcement view of *E*'s impact and a cognitive-restructuring view. The fact that Conditions II and IV do not differ significantly from the respective matched Conditions I and V supports the latter position, but in both cases the condition of partial feedback is closer to the control condition and when *E*'s feedback favors independence, the difference between partial and total reinforcement is almost significant ($P < .10$). There is also evidence that this difference increases as the number of trials increases. Pooling the first 20 trials, the amount of yielding in Conditions IV and V is almost identical. Pooling the second 20 trials, however, the amount of yielding in Condition IV is significantly greater than the amount of yielding in Condition V ($P < .025$). It appears from this that even the reinforcement of independence begins to "extinguish" over a number of trials with no feedback from *E*. This effect may be clearly seen in that portion of Fig. 1 labeled "task set" for Conditions IV and V (see below).

Two conclusions may be drawn from this first experiment: Given orienting instructions which set the *S*s toward individual accuracy, (*a*) feedback in line with the objectively correct alternative inhibits conformity more than feedback supporting an incorrect consensus promotes conformity; (*b*) feedback through the first critical trial has an effect on conformity behavior only when the feedback reinforces independence, and even here the effect tends to wear off in time.

The first conclusion in particular raises some interesting questions. It is rather intriguing to note that this is what one would expect merely if *E* were another *S* added to the group. Thus, Asch (1) found that adding another confederate to a unanimous majority of three had little further effect on the conformity of a naive *S*. However, adding a naive *partner*, or a confederate instructed to respond with the correct answer, significantly decreased the amount of yielding obtained. While it is parsimonious to assume that the present results are effects of at least some of the same causes which determine the pattern of Asch's results, it seems unreasonable to conclude that *E*'s feedback attempts have all of the same effects as an added *S*'s judgmental responses would have. First of all, of course, *E* responds after all *S*s have made their judgments and thus his influence can only be apparent as a function of a learning and generalization process. Secondly, *E* is not making a judgment at all in giving feedback. He is presumably reading the correct alternatives from a master sheet and is thus, except when *S* becomes suspicious, accurate by definition. Because of our conviction that the authoritative feedback of *E* differs from the judgment of an added *S*, a second experiment was designed to explore the effects of feedback when the *S*s have a somewhat different orientation to the task.

EXPERIMENT II

At least two experiments have recently demonstrated the importance of "set" as a variable in conformity experiments. Deutsch and Gerard (5) found that *S*s who were working together to obtain the fewest errors as a group showed greater conformity than when each *S* was merely motivated to minimize his own errors. Thibaut and Strickland (9) provide evidence that, among groups committed to an initial judgment, *S*s who are primarily oriented toward maintaining their status as group members show greater conformity the greater the conformity pressure placed on them. The *S*s who are mainly interested in judgmental accuracy actually conform *less* when the pressure to conform is high than when it is low. In interpreting the findings of both experiments, it may be concluded that there are some situations of individual judgment in a group where it is less important to be accurate than it is to report in line with group consensus, regardless of subjective or objective accuracy. As Deutsch and Gerard have pointed out, in any conformity situation judgments of others in the group provide *information* with regard to objective reality. Situations vary, however, in the extent to which *normative influence* is likely to be operating, or the extent to which *S*s fear the social consequences of nonconformity per se.

In planning the second experiment, it seemed to us that the effects of feedback would differ as a function of the strength of normative versus informational influence. Assuming that informational influence predominated in our first experiment, we saw that feedback favoring independence was more effective than feedback favoring conformity. In a situation where normative influences are more powerful, however, we should expect that feedback favoring conformity would be more effective than feedback favoring independence. The second experiment was designed to test this hypothesis.

Method

The procedure for the second experiment was identical to that of the first except in the following respects:

Design. Since the interest was no longer in comparing partial versus continuous reinforcement, only Conditions I, III, and V were replicated in the second experiment.

Subjects. For reasons of availability, freshmen and sophomore female candidates for the B.S. degree in nursing were employed as *S*s at the standard hourly rate. Because of the nature of the nursing school curriculum and our own methods of recruitment, there was a greater chance that *S*s in a given group would know each other well than in the first experiment, a state of affairs compatible with our attempt to strengthen normative influence. Forty *S*s participated in the experiment proper, and an additional 12 *S*s were later added for purposes of control (see below).

Orientation to the experiment. As in the Thibaut and Strickland (9) experiment, the attempt to strengthen normative influence was composed of two parts: a preliminary routine adapted from procedures (3, 8) for arousing the affiliation motive, and a special set of instructions that the group was competing against other similar groups. These two procedures were presented to the *S*s as unrelated. One experimental assistant introduced the affiliation arousal task and took special pains to make the *S*s feel that this was the first of several sessions in which they would meet as a group. He solicited assurances that the *S*s would cooperate in further experiments involving appraisals of each other (for which they would also be reimbursed). The purpose of this procedure was to make salient the prospect of continuing social evaluation by the group.

After the first experimental assistant had left, the *S*s took positions behind the light panels, and a second experimental assistant gave the apparently unrelated instructions for the judgment task. The instructions were the same as in the first experiment, except that a reward (tickets to a local college musical production) was promised to the five groups making the fewest errors as a group. As in the Deutsch and Gerard experiment, *S*s were told that the group could make as many as four or as few as zero errors on any given trial. In order to make each *S* more aware of the possibility of being rejected for ignoring majority opinion, each *S* was given a mimeographed card designating the geographical positions of A, B, C, and D. From this card, *S* could locate her own position and thus her letter or row. Each *S* was actually given a different card designating her alone as "Subject D"; i.e., the position of D on the card was always identical to the *S*s' geographical position.

Thus, the only changes in the design of the second experiment were (*a*) the omission of the two partial feedback conditions, (*b*) a change which made each *S* think she could be identified by the others, and (*c*) a different emphasis in orientation so that group accuracy rather than individual accuracy was stressed. In addition, an attempt was made to arouse the affiliation motive prior to the judgment task in the second experiment. Following Thibaut and Strickland (9), the orientation of the first experiment will hereafter be called "task set" and that of the second experiment, "group set."

Results and Discussion

The effects of feedback are markedly different as a function of the *S*s' set or orientation to the task. Figure 1 illustrates graphically the joint effects of feedback and set on the average number of yields per *S*. The most striking difference occurs in Condition I, where *E* indicated after each trial that the consensus was correct. In the first experiment, this condition resulted in an average yielding frequency of 9.67, an amount which does not differ significantly from the mean frequency of yielding in the no-feedback condition (III) of 7.50. In the second experiment, Condition I resulted in an average of 13.90 yields per *S* as compared with 4.58 yields for the no-feedback condition. This

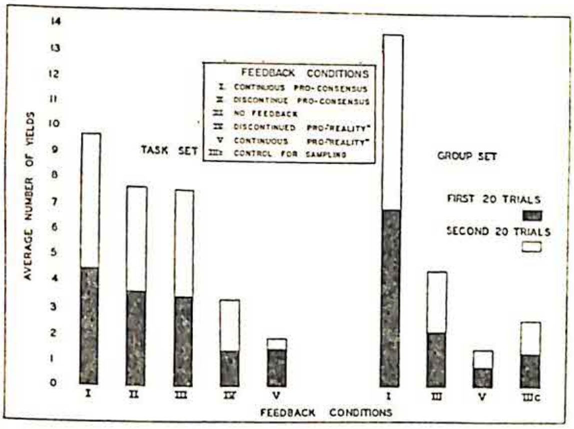

FIG. 1. MEAN YIELDING FREQUENCY BY SET AND FEEDBACK CONDITION

difference is significant both by t test and by median χ^2 test at the .001 level. The difference between these differences (10, p. 159) is also significant at the .001 level.

In Experiment I it was established that feedback that reinforced independence resulted in a highly significant decrement in yielding, as compared with the no-feedback condition. In the present experiment, the decrement is less marked. The difference between the means of Condition III (4.58) and Condition V (1.50) is significant at the .01 level by t test, but the median χ^2 value of this same comparison results in a nonsignificant value.

Interpreting differences between the two experiments is complicated by the fact that the nurses in the second experiment show less yielding under the no-feedback condition than do the students in the first experiment under the same condition. Thus the greater difference between Conditions I and III in the second experiment is partly a function of the elevation of the mean in I and partly a function of the depression of the mean in III. This could mean that our joint attempt to arouse the affiliation motive and stress *group* accuracy actually led to less yielding when no feedback was given, a finding that runs counter to the Thibaut-Strickland (9) and Deutsch-Gerard (5) results. Or it could mean that the samples differed with regard to their tendency to yield without feedback; that the nurses in the second experiment are less conforming than the students in the first experiment.

Although the difference between the two no-feedback conditions is not significant ($P >$ .10), the importance of deciding between these alternative explanations prompted the inclusion of an additional control group. In this group (which we shall call IIIc), nursing students were again used, but the judgment task was administered under conditions identical to the no-feedback group of the first experiment. Individual accuracy was again stressed and no attempt was made to arouse the affiliation motive. As Fig. 1 shows, the results are rather clear-cut. The average frequency of yielding is even less for this control group (IIIc) than for the no-feedback, group-set group. The latter, in turn, is less than the average frequency of yielding for college students under the task set. The only significant difference is that between the nurses and students under task-set conditions. It appears, then, that the samples do differ in their proclivity towards yielding and that group set instructions do raise the level of yielding in the no-feedback condition, if only slightly.

This finding makes the high frequency of yielding in Condition I, group set, even more striking. In this group, six out of 16 *S*s yielded on *every critical trial.* The average *S* yielded on all but two of the 16 critical trials. Clearly the combination of pro-consensus feedback and group-set instructions produces maximum yielding. Thus, one-half of our prediction is confirmed. On the other hand, pro-reality feedback is also somewhat more effective than we had anticipated, and though the difference between Conditions V and III is less marked than under task set, this result may be attributable entirely to the aforementioned decline in the amount of yielding in the nursing group under no-feedback.

DISCUSSION

The present results give support to the statement that the effects of feedback from *E* are contingent on *S*'s perception of the judgment task. Unless *S* feels that an error on her part will jeopardize her standing in the group, she will be relatively unresponsive to feedback favoring group consensus. Findings from the postexperimental questionnaires clearly indicate that *S*s were aware of being influenced by preceding judgments and could estimate with considerable accuracy the extent to which their responses were affected by others. (However, *S*s tended to yield only when they thought the group was correct.) For the most part, then, conformity in the present experiment is quite deliberate—at least some of the time. Under task-set conditions, the motive to conform is probably not of sufficient strength to cause the *S*s to make use of conformity-inducing feedback from *E*. The motive to be accurate and/or to avoid conformity is strong enough to make most *S*s sensitive to feedback favoring independence.

These results differ from those of Crutchfield (4) who found that conformity-inducing feedback greatly enhanced yielding in what appears to have been a task set. This difference can perhaps be explained in terms of differences in the awareness of yielding in the two experiments. Since the *S*s in Crutchfield's experiments were gradually rotated into the crucial position of responding last, it is probable that they were less aware of being influenced in their judgments when reinforcement was applied. This explanation is most tentative, however, and the role of awareness remains an empirical question. Under group-set conditions, the motive to conform is generally greater, but it is important to note that the *S*s are not driven into indiscriminate yielding by their needs to affiliate. The present results show that *S*s who can conform, while at the same time being assured by *E*'s feedback of their accuracy, do so consistently and almost without exception. When one must conform at the expense of accuracy, the tendency to conform is no greater under group set than under task-set instructions. This does not necessarily represent an unresponsiveness to normative influence, since some *S*s may well have resisted consensus in order to secure at least one accuracy point for the group on a given trial.

This consideration points up how difficult it is to create conditions under which normative influence may be divorced from informational, and to create conditions in which predictions with regard to the level of yielding may be nicely coordinated to normative versus informational instructions. As a matter of fact, both in the present experiments and in the Deutsch-Gerard (5) experiment, an essential control condition is absent. *S*s in the normative condition in both experiments are if anything *more* interested in accuracy than *S*s in the informational condition. It may be true that the desire to be accurate stems in part from a wish not to let the group down, but the prize offered for the most accurate groups might, if it were offered on an individual basis, increase the desire to be accurate in any event. In order to check on the effects of increasing the motivation to be accurate, it is necessary to include a control group in which the same amount of reward is offered to all those *S*s who reach a certain criterion of accuracy. If the *S*s under such a condition showed the same level of yielding as *S*s under a group-set condition, then it would be fair to conclude that the present findings under group set and the findings of Deutsch and Gerard for their "group situation" were obtained by increasing the informational influence of the group rather than strengthening normative influence alone. It should be noted that this criticism does not apply to the Thibaut-Strickland study (9), since there the group set involved competition between groups not for accuracy of judgment, but for "effectiveness of cooperation."

In any event, it is clear that feedback from *E* plays a different role from that of the same set of responses made by an additional *S*. While the hypothesis of equivalence was tenable after the first experiment, a number of saving assumptions would be necessary to have it remain tenable after the second. The questionnaire data show that the *S* never questions the accuracy of *E*'s feedback. When his feedback supports consensus, *S* becomes aware that he can increase his own accuracy by going along with the consensus on subsequent trials. Whether he in fact does so appears to depend on the balance of his motives in the situation. The balance can be clearly affected by differences in the instructions which orient *S* to the judgmental task.

SUMMARY

Two related experiments were conducted to explore the effects on conformity behavior of feedback from *E*. The first experiment was primarily concerned with the importance of continuous feedback as compared with early feedback which was then discontinued. The experiment was designed so that this comparison could be made under conditions of feedback reinforcing independence as well as feedback reinforcing conformity. The results showed that early feedback was less effective than continuous feedback, but not significantly so except toward the end of the series of trials, and then only when the feedback reinforces inde-

pendence. While this is suggestive of "extinction," the data are equivocal regarding the importance of reinforcement continuity in learning to conform or not to conform in the experimental situation.

In the first experiment it was also found that feedback in line with consensus has little or no effect, whereas feedback in favor of objective reality drastically reduces the level of conformity behavior. To test the generality of this finding, a second experiment was performed in which the importance of group accuracy was stressed and social evaluation within the group was made salient.

Under these conditions, reinforcement of conformity had a strong effect, and the effect of reinforcing independence was still significant but somewhat reduced. These findings were generally in line with predictions, and they point up the importance of set or orientation in predicting the effects of any independent variable on the tendency of *S*s to conform when placed in a conflict between sensory impression and social consensus.

REFERENCES

1. ASCH, S. E. Effects of group pressure upon the modification and distortion of judgments. In G. E. Swanson, T. M. Newcomb, and E. L. Hartley (Eds.), *Readings in social psychology* (Rev. ed.) New York: Henry Holt, 1952. Pp. 2–11.
2. ASCH, S. E. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychol. Monogr.*, 1956, **70**, No. 9 (Whole No. 416).
3. ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405–410.
4. CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychol.*, 1955, **10**, 191–198.
5. DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 629–636.
6. MOUTON, JANE S., BLAKE, R. R., & OLMSTEAD, J. A. The relationship between frequency of yielding and the disclosure of personal identity. *J. Pers.*, 1956, **24**, 339–347.
7. SCHEIN, E. H. The effect of reward on adult imitative behavior. *J. abnorm. soc. Psychol.*, 1954, **49**, 389–395.
8. SHIPLEY, T. E., & VEROFF, J. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, **43**, 349–356.
9. THIBAUT, J. W., & STRICKLAND, L. H. Psychological set and social conformity. *J. Pers.*, 1956, **25**, 115–129.
10. WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Henry Holt, 1953.

Received July 19, 1958.

THE ROLE OF SOCIAL EXPERIENCE IN THE PRODUCTION OF GASTRIC ULCERS IN HOODED RATS PLACED IN A CONFLICT SITUATION[1]

JOHN J. CONGER, WILLIAM L. SAWREY, AND EUGENE S. TURRELL
School of Medicine, University of Colorado

RECENT studies by Sawrey and Weisz (14, 20) and by the present authors (13) have demonstrated that it is possible to produce gastric ulcers in rats by placing them in a chronic approach-avoidance conflict situation in which approach responses based on hunger drive are in competition with avoidance responses based on electric shock. It has also been shown by the present authors that the factor of psychological conflict per se contributes significantly to ulcer formation in this situation, although hunger and shock also contribute significantly, but only in interaction (13).

In Sawrey and Weisz' pilot study (14), Nebraska hooded rats were used as subjects, and the animals were placed in the conflict situation three at a time. Under these conditions, animals took approximately 30 days to develop gastric ulcers. In the follow-up study by the present authors (13), Sprague-Dawley albino rats were used as subjects, and the animals were placed in the conflict situation alone (because of the demands of the more complex experimental design employed). It was found that it took an average of only 14 days for ulcers to develop in these animals. The ages (100–160 days) and weights of the animals used and the nature of the conflict conditions (i.e., schedule of food and water deprivation and amount of shock) were comparable in both studies.

Thus, it appeared that the obtained differences in susceptibility to ulceration in the two cases might well have been a function of differences either in genetic strain, in social experience (i.e., being placed in the conflict situation alone or in groups), or both. However, since the two types of variables are confounded in the above comparison, it was logically impossible to decide among these alternatives.

The aim of the present study was to investigate the relation of social experience to ulcer formation in a chronic approach-avoidance conflict situation (8), with the variable of strain held constant. It was hypothesized, partly as a result of the above studies and partly on the basis of the findings of other developmental animal studies involving social experience variables, that: 1. Animals placed in the conflict situation alone are less resistant to ulceration than animals tested with other animals present. 2. Animals reared in isolation are less resistant to ulceration than animals reared together. 3. Interaction effects between these two sets of conditions exist. More specifically, the third hypothesis involves the assumption that a change in the animals' social experience (i.e., of aloneness or togetherness) from rearing to testing may (*a*) be stressful in itself and (*b*) may thus increase the overall stressfulness of the ulcer-producing conflict situation, above and beyond any effects of the rearing or testing conditions considered separately.

METHOD

Apparatus

The apparatus used in the present experiment was identical to that employed in the previous studies referred to above (13, 14). Briefly, it consisted of a number of experimental boxes with grid floors formed of brass rods $\frac{3}{32}$ in. in diameter and spaced ½ in. apart. Each box was approximately 36 in. long, 18 in. high, and 12 in. wide. Each contained a food platform at one end and a water bottle at the other. Once the experiment proper was begun, the section of the grid adjacent to the food platform was continuously charged, using a constant current variable resistance shocking apparatus of a type described by Muenzinger and Walz (10). The remaining half of the grid was not charged. Thus, an animal placed in one of the boxes was shocked only if he approached food.

Subjects and Procedure

The *S*s were 40 male hooded rats, born in the investigator's laboratory, and weaned at 21 days. Using the split-litter technique, the experimenters separated the

---

[1] This investigation was supported in part by a research grant (M-930) from the National Institute of Mental Health of the National Institutes of Health, Public Health Service. The present report is based on a paper presented at the annual meeting of the American Psychological Association, New York City, 1957.

TABLE 1

NUMBER OF ULCERS AND INITIAL, PRECONFLICT, AND FINAL WEIGHTS FOR THE FOUR EXPERIMENTAL CONDITIONS

Animal No.	Reared alone Tested-alone				Reared-alone Tested-together				Reared-together Tested-alone				Reared-together Tested-together			
	Wi	Wa	Wf	U	Wi	Wa	Wf	U	Wi	Wa	Wf	U	Wi	Wa	Wf	U
1	280	289	180	1	264	254	173	0	294	309	202	17	290	297	198	1
2	372	372	252	0	402	390	287	0	353	350	224	20	302	309	200	0
3	475	456	342	13	479	467	338	5	385	386	266	22	438	447	320	0
4	392	404	221	21	474	448	341	1	438	429	306	9	452	450	330	0
5	476	466	343	4	483	469	345	8	395	399	265	10	436	434	315	0
6	476	465	323	9	418	398	300	0	391	387	238	15	399	396	280	0
7	363	357	231	5	367	354	238	2	329	329	210	16	390	378	274	3
8	338	339	193	39	330	317	226	0	387	394	227	12	394	401	288	2
9	287	289	235	0	349	339	270	0	311	319	200	27	400	405	308	0
10	320	321	224	4	283	309	222	0	371	390	284	0	333	342	248	2
Mean	377.9	375.8	254.4	9.6	384.9	374.5	274.0	1.6	365.4	369.2	242.2	14.8	383.4	385.9	276.1	.8

Note —Wi = initial, pre-experimental weight; Wa = preconflict weight following five-day adaptation period; Wf = final weight at autopsy; U = total ulcers per animal.

animals from the mother at weaning and divided them into two groups. Animals in the first group were reared in isolation. Animals in the second group were reared with two to three other animals. Both groups received standard laboratory care, including continuous, free access to food and water, except that neither group was handled.

In adulthood (120–180 days in various replications of experiment), half of the animals in each group were placed in the experimental boxes two to three at a time. The other half were placed in this situation alone. Thus, a simple 2 × 2 design with 10 animals in each cell was achieved. At the time they were first placed in the boxes, initial weights were determined for all animals. The animals were then allowed to live in the experimental box for an adaptation period of five days, during which they were not exposed to shock and were permitted continuous access to food and water. At the end of this five-day period, all animals were again weighed (preconflict weights) and returned to their boxes. At this time, the animals were placed on 46½-hour hunger drive, with free access to water, and the shock was turned on in the section of the grid adjacent to the food platform, while the remainder of the grid was not charged. The result was that if an animal approached food, he was shocked. The animals lived in this box throughout the course of the experiment. The shock was turned off for 1½ hours after 46½ hours' deprivation, and the animals were allowed to eat freely during this period.

For each replication of this design, all animals were sacrificed when any animal appeared seriously weakened (14–17 days in various replications). At this time, the numbers of ulcers in the rumen and body of the stomach, considered separately, as well as total number of ulcers, were determined for each animal.

RESULTS

Several different analyses were made. Most importantly, parametric and nonparametric analyses of variance were run for ulcers in the rumen and body of the stomachs separately, and for total ulcers. Because of similar results in all three cases, only results for total number of ulcers are reported here (see Tables 1 and 2). A similar analysis was run for weight loss during the experimental period; also, since weight loss is a potentially important consideration in ulcer formation (16), a covariance type of analysis, as described by Cochran and Cox (5), was employed to determine whether there was a difference in ulcer formation between the four experimental groups, above and beyond any difference that might possibly be attributed to weight loss (Table 2). In the parametric analyses of variance and covariance, a treatments × treatments × randomized block design was employed in order to control for the varying time durations of the 10 replications and possible differences between litters.

As may be seen, the difference in total number of ulcers between the two tested-together and two tested-alone groups is significant at the .001 level, thus supporting the first hypothesis; namely, that animals placed in the conflict situation alone are less resistant to ulceration than animals tested with other animals present. This difference remains significant at the .01 level even after removal of the possible effects of weight loss by covariance. In contrast, reared-alone and reared-together groups do not differ significantly, thus failing to support the second hypothesis, namely, that animals reared in isolation are less resist-

TABLE 2

ANALYSIS OF VARIANCE TO DETERMINE THE EFFECTS OF REARING AND TESTING CONDITIONS UPON THE INCIDENCE OF ULCERATION WITH WEIGHT LOSS CONTROLLED THROUGH COVARIANCE

	df	U²	UW	W²	MS(U)	*F*	Residual			
							df	SS	MS	*F*
Rearing	1	48.4	−42.9	38.0						
Testing	1	1260.0	1567.5	2030.6	1260.0	32.58*	1	442.1	442.1	13.91**
Blocks	9	362.4	1322.2	16829.2						
R × T	1	90.0	49.5	27.2						
B × T	9	348.0	498.5	2651.6	38.6		8	254.2	31.7	
R × B	9	555.6	424.4	2639.2						
R × B × C	9	770.0	368.5	1238.0						
Total	39	3434.4	4187.7	25453.9						

Nonparametric Analysis of Variance (Wilson)[a]

	df	Chi-squared
Testing	1	14.4***
Rearing	1	.4
R × T	1	.4
Total	3	15.2

Note.—U = Ulcers; W = Weight loss.

* Significant at .001 level.

** Significant at .01 level after removal of effects of weight loss by covariance.

*** Significant at .001 level.

[a] Since the underlying assumptions for analysis of variance were not completely met by the data, a nonparametric analysis of variance was also run (21). Since the results of the parametric and nonparametric analyses of variance are highly comparable, it seems improbable that any significant distortion has been introduced into the parametric analysis of variance and, hence, probable that the results of this test and of the subsequent covariance analysis are dependable. This appears especially likely in view of the high levels of significance obtained.

TABLE 3

ANALYSIS OF VARIANCE TO DETERMINE THE EFFECT OF REARING CONDITIONS UPON WEIGHT LOSS DURING THE PRE-EXPERIMENTAL ADAPTATION PERIOD

	df	SS	MS	*F*
Rearing	1	990.0	990.0	12.04*
Testing	1	275.6	275.6	3.35
R × T	1	164.0	164.0	2.00
Within	36	2959.3	82.2	
Total	39	4388.9		

* Significant at .01 level.

ant to ulceration than animals reared together. However, it was the opinion of the investigators that failure to demonstrate a significant effect of the rearing variable may not have been due to an absence of this effect. It appeared possible that animals in the tested-alone groups developed ulcers so quickly that possible differences, attributable to rearing, between the tested-together groups did not have time to emerge, since animals in all four groups were sacrificed when animals in any one group became weakened as a result of ulceration. Obviously, additional research is necessary to further investigate this hypothesis.

As a corollary of the latter hypothesis, however, an analysis of the relation of rearing condition to the possible stressfulness of being placed in a new environment, even in the absence of the ulcer-producing conflict, was attempted. This was done by comparing the weight changes of the four experimental groups during the five-day adaptation period in the experimental boxes, before the factor of conflict was introduced. The results of this comparison are shown in Tables 1 and 3. Animals reared alone have a significantly greater tendency (.01 level) to lose weight during the adaptation period than animals reared together, thus lending some indirect support to the second hypothesis.

Similarly, no significant interaction effects were found, thus failing to support the third hypothesis. However, it was the opinion of the investigators that it was more likely, for reasons analogous to those discussed in relation to the second hypothesis, that a possible interaction effect was being masked by the conditions of the experiment than that interaction did not exist. In order partially to test this assumption, a Mann-Whitney test (6) for the significance of differences in ulceration between the reared-alone-tested-alone and the reared-together-tested-alone groups was run. It was found that the latter group, which had experienced a change from rearing to testing conditions, had considerably greater ulceration. This difference was significant at the .08 level of significance, thus lending some but, of course, not conclusive support for the assumption of an interaction effect.

Furthermore, when reared-together-tested-together animals were compared for ulceration with reared-alone-tested-together animals, the latter—or, again, the "change"—group tended to have greater ulceration, although not significantly so. In this case also, a more sensitive experimental procedure will be required to further explore the possibility of an interaction effect, with respect to ulceration.

DISCUSSION

The expectation, subsequently tentatively confirmed, that reared-together animals would be more susceptible to ulceration than reared-alone animals when placed alone in the conflict situation receives indirect support from a number of studies, at both the animal and human levels. Bernstein (3), for example, in his studies on the retention of a black-white discrimination habit, found that interrupting the handling of rats which previously had been handled since weaning, produced greater disruption of the habit in these animals than in those that had received no previous handling. Similarly, Scott (15) has found that dogs which have been raised in a gregarious environment become severely disturbed emotionally if isolated, even for a few hours.

At the human level, Spitz (17) and others (7) have noted that infants separated from their mothers and deprived of accustomed mother-child contacts tend to develop a variety of unfortunate emotional reactions. They also seem to become unusually susceptible to infection and illness. Furthermore, Spitz noted that the effects of separation were most severe in those cases where favorable mother-child relations had existed previously, i.e., where the change from the child's previous experience was greatest (18).

In all of these situations, as well as in the present ulcer-producing conflict situation, deprivation of previously established social contacts—whether with other animals, with the experimenter, or with the mother—seemed associated with emotional disturbance, increased tension, and reduced resistance to physiological disorders. In the present experiment, it might be argued that, in the case of reared-together-tested-alone animals, heightened stress resulting from unrewarded learned drives previously established in relation to other animals, or from the greater unfamiliarity of the new social stimulus situation, or both, summates with the tension-producing effects of the conflict situation itself to produce greater vulnerability to ulceration.

Of course, it might also have been anticipated that in the tested-together conflict situations, animals reared together would prove more resistant to ulceration than reared-alone animals. In the Bernstein study described above, previously handled animals, when handled during the retention period, made fewer errors than previously nonhandled animals. Several studies (1, 15) suggest that experience with other animals tends to be more stressful for animals reared alone than for animals reared together. As already noted, however, the present study does not provide support for this hypothesis, largely, in the opinion of the investigators, because too few ulcers were produced in either of the tested-together groups (as a result of the experimental conditions) for a significant discrimination to become possible. However, such differences as exist are in the direction of the hypothesis.

If a change in the social stimulus situation from rearing to testing is a factor leading to increased susceptibility to ulceration in the present conflict situation, and if this change is most traumatic in the case where previous relationships with other animals are interrupted during conflict, then it would be anticipated that the greatest difference in number of ulcers would be found between the reared-together-tested-together and reared-together-tested-alone groups. This was, in fact, the case. While the mean number of ulcers in the reared-together-tested-together group was only 0.8, in the reared-together-tested-alone group it was 14.8. The mean numbers of ulcers for the two remaining groups fell between these extremes (1.6 for the reared-alone–tested-together group and 9.6 for the reared-alone–tested-along group). In this connection it is interesting to note that in a study by Benjamin and Bernstein (2), it was found that the mortality rate in previously handled animals after intraperitoneal injection of thiourea was much greater when handling was interrupted during the experimental phase than when it was continued.

Probably the most difficult finding in the present experiment to deal with conceptually is the fact that while, according to the above reasoning, reared-alone–tested-alone animals might be expected to be more resistant to ulceration than reared-alone–tested-together animals, this was not the case. In the present instance, this seems to be a function of the strong ulcer-inhibiting effect of the tested-together condition, as contrasted with the tested-alone condition, regardless of an animal's previous experience of togetherness or aloneness. This effect was, of course, originally predicted by the experimenters, but largely on empirical grounds as a result of previous experimentation. Understanding the reasons for the effect is quite a different matter.

While insufficient evidence exists at present for reaching a decision, three possibilities seem most likely. One of these, and probably the easiest to assume but the most difficult to demonstrate, is that there is a constitutionally determined capacity for the presence of other animals of the same species to have a tension-reducing effect, even in the absence of previous social experience.

Another possibility is that in some way the presence of other animals in the conflict situation has a facilitating effect, by leading to learning which tends to reduce the stressfulness of the experimental conflict situation. Conceivably, this could occur in several ways. For example, at least some of the animals in a group placed in the conflict situation together might learn to become followers through a process of matched-dependent imitation, as described by Miller and Dollard (9). Thus, they might learn that going up to eat is "safe" only when they see another animal eating, and that staying away is necessary to avoid shock when other animals are staying away. Even if this imitative behavior were only partially learned, it would tend to reduce the strength of the conflict for these animals, and, in addition, might well lead to their having fewer actual experiences with electric shock.

In order to investigate this possibility further, it is the authors' intention to follow up the present study with additional replications in which the number of attempted crossings of the electrically charged portion of the grid floor will be recorded separately for each animal under all experimental conditions. In the tested-together groups, high-speed cameras operated by electronic relays will be used in order to identify the member (or members) of a group attempting each crossing. By means of this technique, it will become possible to determine definitely whether tested-together animals tend to make fewer crossings than tested-alone animals. It will also become possible to determine whether there is any tendency, within tested-together groups, for one animal to adopt a dominant role and make an undue proportion of the group's total number of crossings. If the latter should prove to be the case, it would then be important to note whether this "leader" animal had a greater tendency to ulceration than other members of his group. Such information should help to clarify further the specific ways in which the social experience variable of the presence of other animals operates to reduce susceptibility to ulceration in the conflict situation.

It is also possible that more directly physiological factors might be involved. For example, in the present experimental situation, animals were allowed free access to water. Studies by Rasmussen (11) and Bruce (4) indicate that rats in a social situation drink more frequently (although apparently the total amount consumed is no greater) than rats which are alone. If fairly frequent ingestion of water increases the animals' resistance to ulceration by keeping the concentration of stomach acid relatively diluted, fewer ulcers might be expected in the tested-together animals. Unfortunately, amount and rate of water consumption could not be controlled in the present study.

In further research, both water and food consumption will be measured as accurately as possible. It does not seem likely, however, that reduced food consumption on the part of tested-alone rats, even if it should be demonstrated, could adequately account for the greater ulceration in these animals since significant differences persist even after the removal of possible weight-loss effects through covariance. Nevertheless, more adequate information concerning food consumption is desirable, in part because such information would make it possible to study the efficiency of the body's utilization of food consumed under the various experimental conditions of this study.

A third general possibility with respect to the present finding stems from the fact that while so-called reared-alone animals were reared alone from weaning, they were exposed to contact with the mother (and their siblings) from birth until weaning. It is conceivable that interanimal contacts during this period could have gained sufficient reinforcing (i.e., tension-reducing) value, for similar contacts to have reinforcing value in the conflict situation in adulthood. While an interspecies analogy may appear somewhat farfetched, it is at least worth noting that such a hypothesis would be consistent with psychoanalytic formulations regarding the effects of early mother-child relationships in humans. It also appears consistent with the findings obtained in several infrahuman studies (1, 12, 19). It seems to the present authors impossible to express a definite preference for one or another of the above possibilities without further experimental investigation.

Since a distinct rearing effect was not demonstrated in the present experiment (although in the authors' opinion it still seems possible), it seems fruitless to speculate as to a possible basis for any such effect.

At any rate, the present experiment appears to demonstrate that the social experience variable of the presence or absence of other animals can affect the susceptibility of rats to gastric ulcers when placed in a chronic approach-avoidance conflict situation. To this extent, the present findings appear consistent with those obtained in other studies, where relationships with other social beings—whether the mother, an experimenter, or other members of the same species—seem capable of affecting an organism's emotional responsiveness and its resistance to illness and disease.

Summary

The role of social experience variables in the production of gastric ulcers in rats placed in a chronic approach-avoidance conflict situation was investigated. It was hypothesized that: 1. Animals placed in the conflict situation alone are less resistant to ulceration than animals tested together. 2. Animals reared in isolation are less resistant to ulceration than animals reared together. 3. Interactions between the two sets of conditions exist, i.e., that a change in the animal's social experience (from togetherness to aloneness or vice versa) may be stressful in itself and may hence increase the over-all stressfulness of the ulcer-producing conflict situation.

With 40 male hooded rats serving as subjects, it was found that: 1. Animals tested alone had significantly greater ulceration than tested-together animals (.01 level of confidence). 2. Reared-alone and reared-together animals did not differ significantly in ulceration. 3. No significant interaction effects were obtained. 4. However, it was felt that the failure to demonstrate significant support for the latter two hypotheses was probably due primarily to limitations in the sensitivity of the present experimental procedure.

REFERENCES

1. Beach, F. A., & Jaynes, J. The effects of early experience upon the behavior of animals. *Psychol. Bull.*, 1954, 239–263.
2. Bernstein, L. A reply to Scott. *J. abnorm. soc. Psychol.*, 1956, **53**, 141–142.
3. Bernstein L. The effects of variations in handling upon learning and retention. *J. comp. physiol. Psychol.*, 1957, **50**, 162–167.
4. Bruce, R. H. An experimental analysis of social factors affecting the performance of white rats. III. Dominance and cooperation motivated by water and food deprivation. *J. comp. physiol. Psychol.*, 1941, **31**, 395–412.
5. Cochran, W. G., & Cox, G. M. *Experimental designs.* New York: Wiley, 1950.
6. Fix, Evelyn, & Hodges, J. L., Jr. Significance probabilities of the Wilcoxon test. *Ann. math. Statist.*, 1955, **26**, 301–312.
7. Goldfarb, W. Psychological privation in infancy and subsequent adjustment. *Amer. J. Orthopsychiat.*, 1945, **15**, 247–255.
8. Miller, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. I. New York: Ronald, 1954. Pp. 431–465.
9. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.
10. Muenzinger, K. F., & Walz, F. C. An examination of electrical current stabilizing devices for psychological experiments. *J. gen. Psychol.*, 1934, **10**, 447–482.
11. Rasmussen, E. W. Social facilitation in albino rats. *Acta psychol.*, 1939, **4**, 275–294.
12. Ruegamer, W. R., Bernstein, L., & Benjamin, J. D. Growth, food utilization, and thyroid

activity in the albino rat as a function of extra handling. *Science*, 1954, **120,** 184–185.

13. SAWREY, W. L., CONGER, J. J., & TURRELL, E. S. An experimental investigation of the role of psychological factors in the production of gastric ulcers in rats. *J. comp. physiol. Psychol.*, 1956, **49,** 457–461.
14. SAWREY, W. L., & WEISZ, J. D. An experimental method of producing gastric ulcers. *J. comp. physiol. Psychol.*, 1956, **49,** 269–270.
15. SCOTT, J. P. The relative importance of social and hereditary factors in producing disturbances in life adjustment during period of stress in laboratory animals. *Res. Pub. Assoc. Res. nerv. ment. Dis.*, 1950, **29,** 61–71.
16. SELYE, H. *The physiology and pathology of exposure to stress.* Montreal: Acta, 1950.
17. SPITZ, R. A. Psychogenic diseases in infancy: An attempt at their etiologic classification. In A. Freud, H. Hartmann, & E. Kris (Eds.), *The psychoanalytic study of the child.* Vol. I. New York: International Universities Press, 1946. Pp. 255–275.
18. SPITZ, R. A., & WOLF, K. M., Anaclitic depression; an inquiry into the genesis of psychiatric conditions in early childhood. II. In A. Freud, H. Hartmann, & E. Kris (Eds.), *The psychoanalytic study of the child.* Vol. II New York: International Universities Press, 1946. Pp. 313–342.
19. WEINENGER, O. The effects of early experience on behavior and growth characteristics. *J. comp. physiol. Psychol.*, 1956, **49,** 1–9.
20. WEISZ, J. D. The etiology of experimental gastric ulceration. *Psychosom. Med.*, 1957, **19,** 61–73.
21. WILSON, K. V. A distribution free test of the analysis of variance hypotheses. *Psychol. Bull.*, 1956, **53,** 96–101.

Received July 29, 1957.

CASE REPORTS

THE OEDIPUS COMPLEX IN A CASE OF DETERIORATED SCHIZOPHRENIA

JACK ROSBERG
Sherman Oaks, California

AND BERTRAM P. KARON[1]
Princeton University

THE existence in psychotic patients of classical Oedipal problems as described by Freud is unmistakable. The authors feel that any observer who comes in close contact over a prolonged period of time with psychotics cannot fail to see evidence of such fantasies. However, the meaning, as we see it, of such fantasies seems to be somewhat different from the classical interpretation and more in keeping with Freud's 1931 statement, in a different context, that "... it seems that we shall have to retract the universality of the dictum that the Oedipus complex is the nucleus of neurosis ..." (1, p. 253).

The significances which we have found in psychotic patients for such fantasies were most clearly presented to us by an apparently deteriorated and almost unreachable patient. The patient had a long history of slowly developing problems. He had been sent to school after school for disturbed children, each school eventually finding him too difficult to handle and sending him on to a school for the more severely disturbed. Eventually, he entered a conventional outpatient analysis at the age of 17. After a short period, his analyst gave up, saying he was psychotic. He was sent to a psychiatric institution, where he was unsuccessfully treated by standard procedures, including electric shock and insulin treatments. After physical therapies had proved unsuccessful, psychological therapies, including several psychoanalytic approaches, were attempted at several different institutions. These also failed to remit the psychosis.

At this point, the authors, whose own approach to the treatment of psychotics is a modification of direct analysis, began treatment. The patient, a young man of muscular build, was approximately 26 years old at the time. At this point he showed little responsiveness to his environment and had lost all control over his bodily functions. It was necessary in the treatment milieu to care for him in every respect—feed him, clothe him, bathe him, care for all his physical needs. This was complicated by the fact that the patient was given to periodic outbursts of violence. These were accompanied by a relative improvement in contact to the point where he would utter one or two intelligible words. The most frequent of these were "room," "dear," "drink, drink," "the work," and "get out."

There were marked signs of physiological disfunction. The patient's hands were generally cold, his fingernails blue, and his skin showed goose pimples even in the hottest weather. His digestive processes were sluggish as evidenced by the odor of decaying food which emanated from his mouth, due not to a mouth but a stomach condition. It will be noted that all of the physiological disturbances manifested by the patient are components of the normal terror syndrome—vasoconstriction, slowed digestive processes, and lack of control over excretory functions. With respect to the last, however, he would at times manifest an extreme control over his bowels, with no movement occurring for weeks at a time.

Interestingly enough, in his bursts of aggression, he might lash out with equal vehemence at the people in the treatment milieu, at thin air, or even at himself. This was a considerable problem inasmuch as his physical strength had enabled him to break the jaws of two attendants at one institution.

The nights were fraught with terror for this patient; his sleep was light and the people who cared for him were awakened nightly by his terror-stricken screams, which were apparently in response to his hallucinations.

Because of the patient's strength and potential violence, it was necessary to keep him in restraints a good deal of the time.

At the time the authors began working with the patient, the junior author spent eight continuous hours with him. During this period, nothing the patient did or said was intelligible to the therapist; nothing the therapist said seemed to have the slightest impact. As a last resort, the therapist sat the patient in front of a TV set, hoping that the screen would hold the patient's attention, but it did not. The therapist sat down next to the patient, placed his arm around the patient's shoulders and said, "What a nice baby you are." The therapist hoped that the physical contact might reach the patient at some preverbal level. He continued this for a half hour. The patient showed little apparent awareness of even this contact.

---

[1] This paper was completed while Bertram P. Karon was a Public Health Service Research Fellow of the National Institute of Mental Health.

In all future sessions, the two authors worked together. This had several advantages. Physical restraints could be discarded without danger to the therapists; the patient could be induced to move around, which seemed to "loosen" the psychosis; the emotional intensity of the therapy session could be maintained at a fairly high pitch for longer periods of time (by alternating therapists), and the therapist not working at the moment with the patient could pick up cues that were missed by the therapist who was intensely involved with him. Moreover, it is of prime significance that what restraint and control were necessary were imposed directly by the therapists and not by artificial restraints. The import of this will be discussed later in some detail.

It might be argued on the basis of classical techniques that the use of two therapists would lessen the therapeutic relationship by "diluting" the transference. The original object relationship, it might be pointed out, consisted of one person—the mother. However, it is our experience with psychotics that the needs of the patient for mothering are so intense that he readily forms a transference to several people. Moreover, it is questionable what meaning can be assigned to the concept of "diluting" the transference unless a therapeutically usable transference exists. In this patient, although transference needs were present, they were therapeutically unusable because of his massive psychotic defenses. It was only by the combined efforts of both therapists that these massive defenses were penetrated and a workable transference established.

Incidentally, the apparently apathetic, deteriorated, unresponsive appearance of such patients is at least in part an active defense. It is a successful defense insofar as it exhausts any therapist who attempts to breach it. But by the continued persistent efforts of two therapists who alternated in a continued long-term high-pressure session, this patient's apparently almost insuperable defenses began to give way.

Each day the therapists devoted several hours to intensive continued work with the patient. A single question based on his fragmentary utterances, such as "What happened in the room?" might be repeated to the patient innumerable times day after day until the patient seemed to tire and respond. The lack of mechanical restraint and the willingness of the therapists to deal with the patient's aggression directly gave him a feeling both of freedom and of security. He had the assurance that nothing he could do could place him in jeopardy, that he could always depend on the therapists to control those urges in himself of which he was afraid. That this control would be exerted directly by the therapists enhanced the personal relationship and helped to build the therapists into the omnipotent benevolent parent figures which the sick patient typically has never had but which he wishes for so desperately.

The first meaningful material which emerged from this patient (after about a week of such therapy) did not deal directly with the oedipal fantasies, but with a homosexual episode. In response to the question, "What made you crazy?" the patient finally said he was sick because of what he and his brother did. On further questioning, he stated that they had performed fellatio[2] upon each other. The patient then became angry with no apparent cause. The senior author asked him why he was mad at his brother. The patient responded saying that his brother had performed fellatio upon him first (that is, the brother taking the oral role and the patient taking the genital role), instead of letting the patient take the oral role first.

On the following day, the first oedipal fantasy appeared. One of the words which the patient had from time to time muttered aloud was "the room." Again and again the therapists had repeated the question, "What happened in the room?" Finally, the patient responded that he had seen "them making love, being affectionate, knowing each other." In response to further questioning, he said that he was "in the room. The door was stuck." He couldn't get out, and he was scared. When asked who was on top, the man or the woman, the patient said, "The woman."

> [And which one did you wish was dead, the one on the top or the one on the bottom?] The one on the bottom.

Here we see a classical primal scene with oedipal feelings expressed, whether it be reality or fantasy. The child sees the mother and father having intercourse, and he wishes the father were dead.

On the following day we explored this material further. The active defense of the apparent inability to concentrate was demonstrated by the fact that the patient would talk coherently until the threatening material was approached and then he would get dreamy and fragmentary. He finally admitted that he had been in the room and had watched them have intercourse.

> [Who were they?] The big. [The big who?] The parents. [Who was on top?] The lovely wife. [Who was on the bottom?]

---

[2] The discussions of sexual material were, for the most part, carried on in the Anglo-Saxon terms that were most characteristic of the patient's own speech. Less printable than technical language, this form of discourse seemed much more meaningful emotionally and much more effective communicatively.

The patient would not answer this question. After repeated interrogation, the patient muttered something about "the brother," but was unwilling to elucidate.

Tiring of this tack, the therapist returned to asking what happened in the room, and when the patient said he saw them having intercourse, the therapist asked him how he felt about that.

> I have to give them a shot. [Both of them?] One of them. [Which one?] The lovely wife.

He then drifted off. The therapist brought him back again and again to the room situation. On one of these occasions he was asked how he felt and he said, "Western."

> [What does that mean?] The cow.

Again he became incoherent. After several more attempts by the therapist to bring him back to the room situation, the patient suddenly said, "Ouch." He seemed to be holding his testicles with his right hand, which was in his pocket. He then said he wanted his mother to play with it, but she might hurt it.

> [Did they say they'd cut them off?] Yes.

He was then asked who it was that said they would cut off his testicles, and he replied, "She."

> [Did your mother say she would do it?] She didn't mean it.

Here we may note that the castration threat seems to come from the mother and not the father. This is typical of castration fears as we have encountered them in psychotics.

The pattern was much more clearly demonstrated in a later therapy session. After describing an incestuous fantasy concerning his sister, who he said resembled his mother, the patient was plied with classical interpretations. For half an hour or more, the senior author (to test the theory) attempted to get the patient to react to interpretations that his incestuous wishes led to a fear of castration by the father. These repeated and intensive interpretations produced no sign of anxiety or of recognition. Finally, after an unsuccessful leading question, the therapists allowed the patient an opening for describing the mother as the castrator.

> [If your father had a knife, what would he do to you for incest?] I'm not sure.

The junior author interrupted at this point and asked whether it was his mother or his father who would take the knife to him for having intercourse with his sister.

> My mother. [What would she do to you?] Cut it off. [Cut what off?]

The patient said that it was his penis that would be cut off. It was possible at this point for the therapists to gain insight into the dynamics of this fantasy by pursuing their line of inquiry. They asked the patient why his mother would cut off his penis for incest.

> So that she could get a platter regularly.

The patient was asked whether he meant that his mother could use his penis "to feed her whenever she wanted." "That's right, so that she could get a platter all the time."

Thus, the patient told us that the reason he felt that his mother would castrate him was not as punishment for incest, but because she wished to appropriate a source of food.

Intercourse is viewed by the patient as a feeding experience with the penis doing the feeding and the vagina being a mouth (which makes accountable the patient's earlier use of the word "cow" in the primal scene).

Penises, he said, go into vaginas, and "a penis is like a breast. Breasts give milk."

The patient told of another incident in which his sister had "dropped her pants" and showed him her vagina. He said she was trying to seduce him because, "she wanted me to feed her. She was like my mother."

One of the most effective questions employed by the therapists proved to be, "What's the worst thing that ever happened to you?"

On one occasion when this question was asked, the patient responded by saying that his mother had taken "it" off and "put it in the drawer."

> [What?] Her bathing suit.

The patient said that she had sat down on "the desk" and spread her legs and showed him her vagina and that he was scared. "She was always taking things away from me." He continued that he was afraid "she wanted me to feed her" and he couldn't. The patient was asked whether this was an "illusion" or reality, and, after some hesitation, the patient replied: "I think it really happened once."

On another occasion in response to the inquiry, "What's the worst thing that ever happened to you?" the patient replied:

> It hurts in the stomach. [What hurts?] Being hollowed out.

It was clear at this point that the basic problem was one of being fed, filled up, as opposed to having to feed, to give up one's products, to be hollowed out. A demand is felt from the mother that he feed (have intercourse with) her, which conflicts with his own never fulfilled need to be fed. He characterized vaginas as being "sadistic," and once described his mother's vagina as being "like a food chopper."

His reaction is shown in the therapy session in which the therapists brought up the subject of intercourse. The patient was asked whether he preferred having intercourse with an anus or a vagina. He said he preferred anal intercourse. He was then asked whether he preferred intercourse with a man's anus or a woman's anus. He said he preferred a man's anus.

> [Why?] Because you can get a drink. [From what?]

From the penis, the patient said.

> [Where?] (After a pause) In the arm.

This reply puzzled the therapists. Eventually, the therapists got the idea that he meant a magic vagina which he formed with his fingers while he talked. When he was asked if this was the case, the patient said that it was. The therapists decided to put this to a further test at a later time, which they did by having the senior author form a circle with his fingers and ask the patient:

> [What is this?] A zero.

The patient was asked whether it was a vagina, and he replied that it was not. The therapist then made the opening long and narrow, and again asked the patient whether it was a vagina. When the patient again said it was not a vagina, the therapist asked him to make it into a vagina. The patient reached out and reshaped the therapist's fingers until they were a replica of his own hand.

This wish to have a vagina, which was manifested by the magic vagina on his hand, came out clearly at another time when the patient said that he envied his sister. When asked why he envied his sister he replied that he envied her because she had a vagina. When asked why he envied her vagina he said, "Because it is beautiful and the old guy likes it."

On another occasion when he said that he wanted a vagina, he was asked why he wanted it and replied, "So that I can get my groceries through it."

His wish for a vagina implied a wish for castration; not only would he have another "mouth" to be fed through, but the possibility of his being called upon to act as a feeder would be eliminated, or, as he stated, no one could make him urinate in their mouths.

His visual hallucinations consisted of innumerable repetitions of two themes. The first of these consisted of his mother or sister making such "seductive" and terrifying demands that he "feed her." The second consisted of a series of related "room" fantasies. These "room" fantasies were related by the patient in response to the question, "What happened in the room?"

By far the most frequent answer related to the primal scene, which was apparently also the most frequent hallucination. The patient described this primal scene in detail in a therapy session, saying that they (his parents) were standing up and that his mother was bending over (*more ferarum*). He demonstrated the position. He said that he saw his father's penis go in but he didn't see it come out, and "I felt bad."

In a later therapy session, he commented on the primal scene, "I want to go home, but they don't want me."

At times when the patient was asked, "What happened in the room?" he would describe other scenes. In each of them, however, there were two people in the room. Sometimes they were his mother and stranger having intercourse; sometimes they were his father and a stranger having homosexual intercourse; sometimes they were his father and his brother having homosexual intercourse; once it was his mother and his brother without specifying their activities. Although the most frequent event referred to was the mother and father having intercourse, the most disturbing was the session during which he said that the two people in the room were his mother and sister. During this session, the therapists had asked him insistently again and again, "What happened in the room?"

The patient was at first mute, then as the questioning continued unabated, got very angry. He rose from the chair in which he was sitting, but since each of the therapists held one of his arms he was unable to throw a punch. He nevertheless kicked out in front of him (luckily, neither therapist was directly in front of him) with sufficient force to split the table that was in the therapy room. When this burst of violence produced no effect on the therapists (who simply held his arms and repeated the question), the patient sat down and began to cry. The therapists repeated their question, "Who was in the room?" and the patient sobbed that his oldest sister was in the room on the bed with his mother. In the further elucidation of this, the patient explained that his mother was feeding his sister. As the patient talked, it became clear that, inasmuch as he viewed intercourse—both heterosexual and homosexual—as a feeding situation, all of the "room" fantasies, including the primal scene, dealt with a single theme: Two people are in the room, including at least one from whom I expect food. Someone else is being fed and I am not wanted.

In summary, a severely regressed case of schizophrenia produced classical oedipal material, including a classical primal scene, during the process of a developing therapeutic relationship. Upon further elucidation, the material seemed to show dynamics which differ somewhat from the classical interpretations of the oedipal situation.

The patient interprets intercourse as an oral situation, equating the vagina with the mouth and the penis with the breast. His own needs to be fed and mothered were not met by the mother; rather he sees her as making demands on him to "feed" her. Thus incest is seen as a feeding situation whereby he is forced to feed his mother against his will. Her demands seem so all-engulfing as to swallow up the penis itself, i.e., the mother and not the father is seen as the castrator. But underlying the castration fear is not only these insistently felt demands from the mother but, also, the wish to be castrated so that such demands can no longer be made. In place of the organ through which he can be drained, he wishes for a vagina through which he can be fed. Such a wish means turning from females to males for "feeding." The primal scenes seem to have been experienced as the mother's devouring needs being met by the father, and neither of them caring about the needs of the child.

REFERENCE

1. FREUD, S. Female sexuality. *Collected Papers. Vol. V* London: Hogarth, 1950.

Received August 23, 1957.

A NEW THEORY OF SCHIZOPHRENIA

ANONYMOUS

FOREWORD.—This remarkable paper was written some two years ago by a then 34-year-old man residing on a closed ward in one of our large VA hospitals with a diagnosis of "paranoid schizophrenia." Soon after the original manuscript came into my hands (through a former student), I had it mimeographed and, in the interim, as occasion arose, have obtained reactions thereto from several clinical psychologists, psychiatrists, and theologians. The evaluations fall into two categories: some hold that the paper is just what it purports to be, an original and highly reasonable *theory* of schizophrenia—while others insist that it is a classical exhibit of the *disease itself*.

The present judgment of the author, now no longer hospitalized, is that his paper, as it stands, is probably valid as far as it goes but needs to be elaborated and extended in certain ways. It is hoped that his further thinking in this connection can be separately published at a later date. In the meantime, readers will be free to form their own opinions as to whether the production presented herewith is an essentially valid, albeit perhaps partial, theory of schizophrenia—or just a "phenomenological perspective."

In its original version this paper was too long for journal publication and has therefore been shortened and somewhat rearranged (in collaboration with the author), in order to improve sequence of argument and eliminate repetition. Three initialed footnotes and one bracketed insert have been added where confusion might otherwise result.

Here it need only be added that the author of the following essay has a history of personal waywardness and perversion (hence the anonymity) which could easily support his original thesis. He regards the present document as largely an expression of unconscious processes, rather than as something that he carefully reasoned out. It was written, he says, "while I was still following signals and colors; some of it was almost automatic writing." The author gives additional information about himself at the end of the paper.—O. HOBART MOWRER, *University of Illinois*.

I WISH to offer a complicated set of hypotheses which form a systematic whole in keeping with but extending beyond certain common-sense and professional Freudian notions.

There have been many recent, much-publicized scientific "breakthroughs" in physics, metallurgy, electronics, and physical medicine. But one hears only that mental illness will soon be conquered. Judging from current newspaper and magazine articles about mental illness, there are two main trends, simultaneous yet diverging, in the professional thinking of psychiatrists and psychologists; one, an increasing acceptance of Freudian psychology with an attempt to apply it to psychoses—this trend is winning over psychiatric diehards who held to the pre-Freudian emphasis on empirical diagnoses, manipulation of the environment, and custodial care, and who had no sound theoretical rationale which would explain observed dynamisms and guide therapy and research; two, a wholly opposed trend of psychiatrists and others who feel that they are perhaps post-Freudian in that they more or less accept Freudian psychology but sense, where they cannot prove, its bankruptcy in accounting for psychoses although it does explain neurosis and is hauntingly fertile with respect to psychosis. This group feels a need for new basic concepts, breakthroughs in theory, and many of them are pinning their hopes on the possibility of physiological and pharmaceutical cures being discovered.

Somewhat in line with both of these groups, I believe that the body of present-day Freudian-type theory of psychosis is nearly all sound, but that it lacks a head—which I have hoped to supply. Freudian theory well accounts for some of the dynamisms in psychosis, but it falls short of organic wholeness and intellectual satisfactoriness.

THE REIGN OF SECRET TERROR

I propose that the motive force of schizophrenic reactions is *fear*, just as fear motivates, according to Freud, neurotic mechanisms—but with these differences: in the case of schizophrenia, the chronic fear is more properly called terror, or concealed panic, being of the greatest intensity; and second, as is not the case in neurosis, the fear is conscious; third, the fear itself is concealed from other people, the motive of the concealment being fear. In neurosis, a sexual or hostile drive, pointing to the future, is defended against. In schizophrenia, by my view, detection by others of a guilty deed, the detection pointing to the past, is defended against.[1]

The following are some semantic, literary, and proverbial evidences of the centrality of fear in psychoses (such types of evidence being respectable since Freud). If schizophrenics appear, through dumb show, physiognomy, tone of voice, or verbal connotation, to lack presence of mind, it is because fear has "scared them out of their wits"; if they appear harried and harrowed, it is because "fear is a cruel master"; if they appear constrained, it is because they are "petrified" with fear; if they appear unfriendly it is because, just as "perfect love casteth out fear," so, conversely, fear casteth out perfect love; if they appear to hate it is because, as "hate breeds fear," so fear breeds hate; if they appear to lack vitality it is because they are "scared to death"; if they appear defeated it is because "the coward dies a thousand deaths"; and lastly, if they lost bladder, bowel, or other "physiological" control, it is because, if an instructive scurrility may be used, they are "scared --------."

---

[1] The author, in accepting the Freudian theory of neurosis, was apparently unaware that others have rejected it in favor of a view which is much more in accord with his own perception of the dynamics of schizophrenia (see below).—O. H. M.

The Theory Stated

My hypothesis may be called the Dick Tracy theory loosely in honor of that familiar fictional, human bloodhound of crime.

1. Motivated in the very first place by fear, the schizophrenic psychoses originate in a *break with sincerity*, and not in the classically assumed "break with reality." The patient's social appetite (an instinctive drive in primates, I believe), including love and respect for persons and society, is consciously anticathected or forsaken and ultimately repressed with the passage of time, since full satisfaction of sociality entails, more or less, communicative honesty, faith, and intimacy. Also, the tension set up in interpersonal intimacy by the withholding of emotionally important (although perhaps logically irrelevant) information causes unbearable pain. This repression of sociality accounts for the well-known "indifference" of schizophrenics. But if safety can be achieved by means of "perjury" alone without great discomfort, then no further defenses are adopted. Perjury is here defined as avoiding telling the whole truth and nothing but the truth. If, for many possible reasons, perjury is not an adequate and comfortable guarantee of safety, as it usually is not, the "cutting" of social contacts is progressively pursued—all in the interest of safety in respect to avoiding possible punishment. Suppression and repression of the social appetite or instinct is thus central in schizophrenia. I believe that repression of sexual and hostile drives is *not* primary in schizophrenia, although it is secondary, as will be explained further on.

Schizophrenia is the cultivation of a lie. A lie is "proved" to be the "truth." The real truth is that the schizophrenic is responsibly guilty of some crucial misdeeds. Simple schizophrenics, hebephrenics, and catatonics "prove"—by words and actions which are louder than words—that they are not responsible. Paranoiacs specialize in "proving" that they are not guilty, could not be guilty because of natural loftiness. But even most paranoiacs avoid responsibility like the plague. Those who accept responsibility see to it that they fail in a big way—like Adolf Hitler.

I believe schizophrenics tend to adopt any means which promise to interfere with the examiner's understanding of them—despite all appearances to the contrary. In order to conceal misdeeds *for the long run*, with minimum disadvantage, very subtle means of uncooperation must be employed. Psychoeconomic factors varying among individuals, account for the specific set of stratagems adopted in any one case.

In my view, a large amount of the damage to the schizophrenic's self-esteem results from his contemplation of his own vicious insincerity, which damage is more an effect than a cause of his disease. His unethical defense mechanisms cause him deep shame and fear of loss of others' esteem. In addition, the primary deeds—whose exposure and punishment are avoided by the disease—are shameful. Insincerity involves using people always *as means, never as ends*. Kant singled out this principle as a natural (instinctive?) moral principle. It is implicit in Christ's Golden Rule. Failure or unfulfillment of the drive for achievement is more an effect than a cause of the disease schizophrenia.

2. A primary danger which is defended against is *detection* by other people of the patient's *antisocial deeds*, past or present (varying from, say, murder and treason to petty emotional dishonesty and masturbation), of which the patient remains conscious throughout any degrees of progression of severity of his disease. A second primary danger defended against is the patient's own guilty conscience, which self-demands his own punishment. Also, the severing of "genuine" social ties is a kind of treason against the human race, and itself brings secondary guilt. Thus, guilt which is always present in schizophrenia, complementing the danger of externally threatening punishment, explains Clifford Beers' discovery that in psychosis there is a serious impairment of self-esteem.

More broadly speaking, schizophrenia shares with all functional mental illness the ultimate danger of punishment meted out by men, demigods, or gods. Common punishments feared are the being deprived of love of kith or kin, loss of social status, financial security, etc., and especially in the case of schizophrenia the more violent punishments such as being abominated by kith or kin, bodily mutilation, imprisonment, lynching, execution.

This abandonment of social ties and good feeling, in the interest of personal safety, is sometimes starkly simple, as in mutism, but is usually supplemented by the development of "phony" social behavior, that is to say, designedly cryptic or misleading expressions of interests, sentiments, opinions; designedly unfriendly "friendliness"; asking only questions to which the answers are already known; the limitation of speech to severely precensored statements; the limitation of conduct to carefully self-criticized, self-rehearsed strategems, etc. The patient has aggressed, ultimately in self-defense, by means of an undeclared, passive, preventive "war" against his fellow men, and in the interest of preventing defeat (positive victory is soon sensed to be hopeless attainment) most of his knowledge and sentiments, and indeed his spontaneous behavioral tendencies, have been classified "top secret." Whatever words he actually uses are employed, thus, as self-defensive weapons. My Dick Tracy theory offers the explanation that high reaction time results from "biding time" in the service of caution. Time is needed in order to creatively choose a "phony" (not spontaneous, not honest, not satisfying—except with respect to fear) response that will maximally conceal that patient's mental life and thus insure safety.

A semi-instinctive deceptive stratagem leads schizophrenics, like pursued rabbits, to "zigzag," thus baffling the pursuer's expectations. Unpredictability is a stock-in-trade fetish of schizophrenics. The proximate goal is to avoid being understood. The ultimate goal is to avoid punishment. They "nonwant" punishment.

But there is another, diametrically opposed stratagem used, I believe, more often by schizophrenics in extreme danger, and by the least intelligent schizophrenics. This stratagem is to follow a steady course when the pursuer expects variability. Also catatonics tend to use this stratagem, repetition being related to repressed sadism. This is an "unenlightened" stratagem, since it is almost frankly concealing in purpose.

3. Only *after* this more or less conscious process of social insincerity and disaffection has become well established does the psychosis proper begin. The merely insincere state, and up to the point of considerable withdrawal from intimate social contacts, is, precisely speaking, preschizophrenic, and is compatible with fairly good adjustment to society. In this state is found the classical "overideation," which is nothing more than a desperate attempt to find means of insuring safety and of retrieving the happiness which existed prior to the occurrence of the unforeseen, unchosen, unfortunate crisis in the patient's life relating to detection and punishment. Prepsychotic schizophrenia is, I believe, related to "psychopathic personality" (a kind of arrested stage of preschizophrenia), some cases of alcoholism, and even what may sometimes be diagnosed as "obsessional neurosis" or "anxiety neuroses."

If my Dick Tracy theory is correct, there is an empirical check. "Basket cases," amputees, permanent wheelchair cases, those permanently in iron lungs, etc., should have an incidence of schizophrenia low to the point of nonexistence. They would never stand trial. Punishment is out of the question. Thus they have no present motive for schizophrenia. Also this particular case throws into differential relief the question of other causes of schizophrenia. For example, a "basket case" must have an unfulfilled drive for achievement, presumably his self-esteem would be low. But I predict infinitesimally few such cases will ever be found to have schizophrenia.

UNATTRACTIVE FEATURES OF THE THEORY

My theory that the primary cause of schizophrenia is not repression of sexual or aggressive drives, nor has a physical source, I call the Dick Tracy theory because this title, at least to Americans, quickly conveys the core concept in a catching and arresting way. It is not meant to be merely facetious.

My hypothesis diverges from orthodox theory, I believe, in insisting that a large part of the schizophrenic's fear is not a symptom and is not repressed, but is the main pathogenic force and is accessible to consciousness and even verbalization were the psychotic to violate his more or less conscious policy of dissembling.

There are understandable reasons why others may not have developed this view sooner. The theorist is naturally made to appear somewhat hardhearted and accusatory. The families and others related to the schizophrenics would naturally resent the implication of crime. And present-day laws and customs are not so arranged as to make the professional handling of old, untried, and ambiguously actionable crimes a pleasant task. I am not familiar with the laws relating to privileged communication to certain classes of professional people. However, it is easy to imagine and believe that physicians and psychologists would not enjoy making a career of being accessories after the fact or having their work open to such interpretation in the sociological "climate" afforded by present-day attitudes. Nor would professional healers in general enjoy arranging for prosecution of patients.

In other words, this theory, even more so than Freud's sexual emphasis, by its very nature would stand to be unpopular all around, both to the patients and the healers and the social group which surrounds them both. In fact, were it not that the truth promises to be practically invaluable in the long run, I would hope that my theory is not true.

As Freud found Shakespeare's Hamlet to be representative of neurosis, I take Shakespeare's Lady Macbeth to typify schizophrenic psychoses. The motto of the schizophrenic might well be, "Out, damned spot!" and that of the therapist working with schizophrenics, "Find the crime!"

SCHIZOPHRENIA AND CRIMINALITY COMPARED

But if my theory is true, then why are not ordinary criminals such as fill our prisons, psychotic? There are, I believe, a number of points of difference between garden-variety criminals and run-of-the-mill schizophrenics:

1. The typical criminal lacks a strong conscience, which is to say, he does not hate himself

nor self-demand his own apprehension and punishment as requisite to continuing self-esteem, in reaction to his own misdeeds.

2. Society's punishment is apt to be viewed without much fear by ordinary criminals, for several reasons. Typically they are already backward in competition for social, financial, and occupation status. Thus they do not have much to lose by, say, imprisonment. Further, due to their formative experiences, they tend to be realistic about punishment, and so think of it without fantastic dread. Prison may be commonly regarded as a kind of school for further crime, offering self-advancement. And prison offers social and homosexual pleasures and "being in one's own social class" to the typical, conscienceless criminal. Whereas, alas, the typical schizophrenic is a person whose formative experiences have taught him to think of himself as "respectable," "good," to consider companionship with criminals to be a horrible stigma, and there is a tendency for him to have had little realistic experience of punishment *for specific deeds*, and the punishment he has received is apt to be grossly unjust. Typically the schizophrenic has been hated and punished for "no reason," and his natural logic leads him thus to reason that if people hate him and punish him for no just cause, how abominably he stands to be punished, now that people *do* have a just cause, might be inconceivable and unbearable. His upbringing and training in the "school of life" lead him to regard punishment with horrendous, fantastic fear.

3. Criminals in prison have been *already caught* and are thus in process of receiving whatever punishment is due them. Thus, no defenses against detection are any longer necessary, and the prison environment is one in which crimes are actually bragged about. Socrates said that the most unfortunate fate which can befall a person is the fate of the unpunished criminal.

4. The very fact of their apprehension strongly suggests that these criminals never did adopt a policy of complete insincerity and repression of the social instinct, but chose, rather than to undergo a complete personality transformation, to take some risk of being caught—and "luck" was not with them. Also the criminal nature of their misdeeds, such that they are pronounced punishable by law, usually was not much in doubt in the mind of the typical criminal, who has some sophistication in regard to legal and criminal matters as they concern himself.

5. Typically, criminals have premeditated their crimes and deliberately taken account of probable consequences. Whereas, I believe, the typical schizophrenic blindly and naively by psychological pressures or external circumstances has been largely forced into his misdeeds, and the full implications of his misdeeds do not occur to him until considerably after their commission.

But then, why do *some* prisoners develop schizophrenia? Is not this fact inconsistent with the Dick Tracy hypothesis? I believe that in these cases the prisoner often is afraid of being punished by his fellow convicts, say, being beaten or murdered or socially ostracized. In other instances, the prisoner has remaining, outstanding, actionable crimes which if detected would lead to further new punishment by the authorities. In Ganser's syndrome, positive gains—such as escaping from painful labor; escaping from deathly boredom resulting from years of a stereotyped, confined life; a desire to be transferred to a location where his best friends are, etc.—add a motive force to an already developing schizophrenia. Thus fear is seen to remain central in schizophrenia.

Only in general do typical criminals differ from schizophrenics. Both are "against" society, with the criminal on the offensive, the schizophrenic on the defensive. One is scrupulous, the other unscrupulous, etc. But scrupulous and unscrupulous attitudes on the same subject can coexist in the same person.

I offer that the criminally insane are concealing misdeeds *other than* the one which caused them to be designated "criminal."

ON CRIME AND PUNISHMENT

I hypothesize that the offenses behind schizophrenia definitely tend to be nonviolent, or with a minimum of violence. The most common offense, perhaps, is nonviolent or cooperative crimes: incest, perversions, adultery. Other common crimes, sins, and "disgraces" are: fraud, embezzlement, misrepresentation, criminal negligence, conspiracy, sneak thievery, arson, malicious malfeasance, sexual intercourse with those of another race, getting married without love, abortion, blackmail, perjury, fornication, masturbation, pedaphiliacal assault, infidelity to spouse, family, employer, government, etc., petty sabotage, membership in secret societies, etc., mental cruelty, apostacy from a strict faith like the Catholic Church, blasphemy, sacrilege, use of narcotics, etc.—passive, "easy" offenses such as opportunity provides amateur wrongdoers and "respectable" people, rather than violent crimes requiring guns, special tools, gangs, etc.

The dishonesty of the disease itself adds a secondary, symptomatic guilt. Additional guilt often arises from offenses committed *after* the onset of the disease. But in therapy, finding the initial, causative guilt and fear is important.

I believe many who later become schizophrenics were childhood criminals. Their great unsolved problem of childhood was now hedonistically to get rid of guilt feelings. Later, adolescent or young adult crimes had as partial motive the attempt to solve in the present, in miniature, by means of a "model," childhood problems of guilt. I believe the brain tends to cause an "acting out" of prolonged excessive pressures. Thus the immediate wrongdoing responsible for schizophrenic breakdown often may have had itself a pathological motive akin to the neurotic need for acting out.

Unfortunately, our society lacks pleasurable means of "getting rid" of guilt feelings, even childish or noncriminal guilt feelings. Seeking punishment is "suicidal."

My theory, as I said before, is appalling, repugnant, not easy to be "swallowed."

But fortunately, in a large number if not in most cases, I believe, the misdeed behind schizophrenia is not very formidable due to a principle which may be called the Baumes' law principle, after the New York State Baumes' law which takes account of repetition of certain classes of offenses. I hypothesize that fear of punishment is additive or cumulative synergystically so that the whole is greater than the sum of its parts, in accordance with Gestalt theory. This principle works especially when uncriticized by reason, and lends itself admirably to secrets, panic, and haste. There is a natural tendency to regard two misdemeanors, as it were, as being one felony. Ten petty misdeeds make one serious crime. In cases where this principle is notably at work in causing the fear underlying schizophrenia, the fear is unrealistic so far as a naturalistic philosophy can see. Ordinarily, for adults, there exists no agency, unless it be God Himself, which actually metes out total punishment (and then some, Gestaltwise) for all misdeeds. The schizophrenic's "crime" often or typically is a compound of a number of petty crimes or misdeeds; thus his "crime" is unquestionably not punishable, and friends, relatives, doctors, and society can breathe freely. The schizophrenic often has in his background a number of offenses punishable, respectively, by a number of separate authorities and persons. In many cases the feared agencies of punishment are inoperative in the present, due to statutes of limitations, insufficient evidence, unwillingness to prosecute, jurisdictional limitations; the death, absence, or impotence of private pesons whose hatred might have been incurred, etc. Therefore, in many cases, the essential fear causing the psychosis is not in the present (and perhaps never was) "realistic" or true to the facts. To the extent that this is so, the schizophrenic has made a serious judgmental *mistake*, analogous to the "bad bargain" Freud found to have been made in the etiology of neurosis. But typically, the schizophrenic's formative experiences have made him naive with regard to the actual legal nature of crime and punishment, as compared with the typical criminal's high-powered sophistication in these matters.

There is another principle at work in the triggering of the mechanisms of schizophrenia. The principle here offered may be called the "Russian roulette" principle, which, I believe, describes a natural tendency of the human mind, especially notable in uncritical thinking. So the schizophrenic has commonly fallen prey not only to the Baumes' law "error" in thinking but also to this second "error," which is that where there exists a small chance of occurrence of an unpleasant result, this probability is assumed to be a certainty. The more unpleasant the feared result, the smaller the probability which is deemed a certainty for practical purposes. When there is one chance in a million that you will be killed on an airplane trip, say, you do not deign to worry about it. But when the chances seem to be, say, three in five, you make out your will. This principle lends itself aptly to the psychogenesis of schizophrenia. A deed which is open to interpretation as a punishable crime is seldom *certainly* punishable. But a small degree of likelihood of grave punishment, brings about the assumption of certain punishment as threatening. And the price of avoiding otherwise certain punishment seems easy to pay; namely, the price of dishonesty and disaffection in interpersonal relationships. But the price is deceptively low, since a very grave and stubborn disease often results therefrom.

Too, those who develop schizophrenia often have gruesomely unhappy social experiences in their personal histories. These painful experiences encourage a morbid pessimism with regard to expected mercy in cases of doubtful outcome. Any gloomy experiences add to the promptitude with which "the worst" is anticipated.

In a diagrammatic sense schizophrenia is the converse of physical diseases caused by microorganisms. In the latter, a multiplicity of symptoms result from a simple cause, namely, the invasion of the body by a certain class of microorganisms. In schizophrenia, a multiplicity of distinguishable and partially independent factors cause what may be considered a rather simple outcome—radical restriction of social behavior and the development of queer thought processes. For example, a poor knowledge of law, an unhappy childhood, a pessimistic philosophy, etc. are among the multiple factors. In fact, the range is probably endless of the factors which can contribute to schizophrenia, in contrast to the neat simplicity of the germs responsible for a certain type of pneumonia.

But actually, on closer view, there are distinguishable sets of symptoms classically recognized. How does my theory account for them?

The Theory and Types of Schizophrenia

I have already explained the mechanisms common to all types of schizophrenia. Now I would like to categorize it more precisely. The mechanism of insincerity, followed by suppression and repression of social appetite, ultimately draws into suppression and repression some or all of the *nonlibidinal drives*, in the interest of insuring safety from detection. In addition to the social drive, the exploratory drive (curiosity), the drive to general activity, the maternal instinct, etc., including the drives discovered in animals by modern experimental psychology, are suppressed and perhaps in time repressed. These mechanisms, I believe, point to a single unitary mechanism which is instinctive and is in the repertory of potential behavior of every normal person. This mechanism is akin to the "playing dead" of certain animals. It may be called the mental playing-dead mechanism. It is partially irresistibly automatic, in agreement with Freud's observation that in psychosis the ego originally assents.

Now, the impairment of the nonlibidinal drives prevents the full satisfaction of the libidinal drives. Moreover, since the libidinal drives lead to intimacy and communication, they are deemed direct hindrances to concealment, and are defended against. Thus the developing psychotic finds himself unfortunately in a state of libidinal "starvation."

It is at this point that a secondary process develops, which makes the incipient case a full-fledged psychotic. The libido automatically and fully unconsciously breaks out of its cruel confinement in the form of painful symptoms. According to which pregenital drive was strongest before the repression of and regression from the dominant libidinal drive prior to the onset of the disease (usually genital heterosexuality), these new, unlooked-for symptoms take special form.

If the patient's strongest pregenital subordinate drive in the state of mental health was passive homosexuality, the result is paranoid-type schizophrenia (in accordance with Freud's theory), in which the very things most feared, detection and punishment, are automatically sexualized and sought, causing horrible exacerbation of fear.[2] Since fear is the prime cause in the first place, and it has been increased, the patient becomes that much sicker. The increased sickness caused a further diminution of the mite of libidinal satisfaction, resulting in increased urgency of the need to be caught (and perhaps raped). And soon, secondary, then tertiary defenses are set up. This vicious circle accounts for the observed fact that paranoid mechanisms are usually progressive.

Where the strongest pregenital drive is aggressive homosexuality, I offer the coordination that hebephrenia is the result. This combination of aggressive homosexuality with the antidetection or mental playing-dead mechanism results in a secondary defense against attaining the affectionate, intimate objective, the conquered beloved. The patient makes himself a maximally ridiculous and objectionable suitor.

Where the strongest secondary libidinal drive was sadomasochism, the result is catatonia. Since actual injury to other people would tend to increase the primary psychogenic danger of punishment, and since, as is not the case in paranoid type, there is no sexualization of "being detected," the patient engages in mental and muscular self-torture in which he accuses, coerces, and threatens himself, assigns himself hated tasks to which he adds cruel fillips of agony, such as repetition. As some psychologists have observed, this is the most painful type of psychosis. In such agonized desperation, control is sometimes lost (the schizophrenic, in terms of his own sickness, "loses his head," since the whole point of the sickness is to avoid punishment) and "violent" cases result.

I think that perhaps there is another type of disease which is sometimes lumped in the category of catatonia, but which should have a different name. This is the shock state resulting from extreme trauma such as being faced with actual mutilation in combat. Here the entire ego seems to be repressed, not to avoid any detection, but merely to avoid the anguish of tremendous fear. Here the regression is rapid to infantile orality. Freud believed psychoses to originate in repression of the ego. I suspect this account applies mostly to this particular type of disease, which might be called shock schizophrenia or shock psychosis.

So-called simple schizophrenia results where formative experiences, style of life, etc., combine with environmental situations in which casual, promiscuous, social, and sexual relationships are easily obtainable and sunderable, and where the concealed misdeed and feared punishment are of a comparatively less serious nature. Promiscuity affords an adequate defense against detection. If my hyopthesis is correct, simple schizophrenics, if incarcerated in a constant environment, would tend to develop one of the other three, more serious, forms of schizophrenia; but in "going to pot" and "bumming around" there is adequate defense to make unnecessary repression of the social instinct and the primary libidinal drive. Thus there is no distorted outbreak of secondary symptoms. But all this is a matter of degree. In cases where the primary libidinal drive is partially repressed, mild perverted tendencies result which the patient deems can be indulged with safety.

---

[2] At this point, as elsewhere, the author attempts to harmonize his own independent and original observations with certain Freudian presuppositions. The result is not always a felicitous one. Here the author is deferring to the view of paranoia put forward by Freud in the famous case of Dr. Schreber (*Collected Papers, Vol. III*) and very widely regarded, in psychiatric circles, as fully validated. Recently, however, Walters has soberly re-examined the clinical evidence, on the basis of which he was forced to conclude that "a very large discrepancy has been demonstrated between the generally accepted relationship and the quantitative studies of paranoia and homosexuality" and that it was "the prestige of Freud's backing"—rather than the empirical facts—which "has continued to give his hypothesis a dominating role in the psychodynamic interpretation of paranoia" (p. 337—Walters, O. S. A methodological critique of Freud's Schreber analysis. *Psychoanal. Rev.*, 1955, **42**, 321–342.).—O. H. M.

It seems fairly likely that "psychopathic personality" may be a mild, arrested form of simple schizophrenia. This would accord with the observations of many that psychopathic personalities when examined under a microscope, as it were, turn out to have one or more types of schizophrenia. Also, some patients diagnosed as neurotic "criminals out of a sense of guilt" may be mild simple schizophrenics, since, according to my theory, what may loosely be called guilt is the specific virus causing all functional psychoses. I think, also, that simple schizophrenia may be nothing more than a congealed, arrested form of preschizophrenia. Thus a number of different diagnostic categories may be allotropic, that is, environmentally stamped forms of mild schizophrenia temporarily or permanently arrested in its early stages.

MANIC-DEPRESSIVE PSYCHOSIS AND ALCOHOLISM

[This section, though brief, is deleted in its entirety because it adds nothing to the central thesis of the paper as a whole and distinguishes, rather categorically and arbitrarily, between the dynamics of depression and schizophrenia. The author, after attributing the latter to real guilt and fear of detection, overlooks the salience of self-punishment in depression and instead stresses a loss-of-love theme. In conversation he has indicated that this interpretation was based upon the observation that—in at least two cases known to him—depression seemed to follow immediately upon the termination of a "love" (illicit) relationship. A more plausible supposition would be that the disruption of such a relationship is simply an early manifestation of an approaching "attack" of bad conscience (depression).—O. H. M.]

SCHIZOPHRENIA, A SOCIAL DISEASE

Schizophrenia is entirely, in its origin, a social disease. It implies a social group. Indeed it is a type of "social" reaction to a punitive society. Solitude in itself is incapable of producing schizophrenia. That solitary confinement in prison does produce schizophrenia can be explained otherwise. "Solitary confinement" is yet a state surrounded by a group into which the prisoner will again emerge. Such solitude dangerously intensifies, through its being dammed up, the social instinct. The need to communicate becomes urgent. Second, the prime defense against being caught, insincerity in social relations, is rendered useless in prolonged solitude. The subject becomes seduced away from his usual defenses and is under pressure to reveal his important thought processes. Recognition of the danger therein causes the subject to adopt extreme, emergency insincerity and distortion. Only prisoners with self-dangerous secrets to reveal develop schizophrenia in solitary confinement, if my hypothesis is correct.

This "social-reaction" nature of schizophrenia makes unlikely any "physical" cure. The real, enduring cure requires the complete establishment of "moral" integrity with regard to our society—that is, to a certain "normal" degree of integrity, since society itself is normally somewhat false-hearted.

Great fear generates, by a natural tendency of the human mind, a corresponding hatred, the purpose of which is self-defense. But, in self-defense, the hatred itself must be hidden—given the nature of our society. Since fear is the outstanding characteristic of schizophrenics, hatred runs a close second. Repressed hatred (which has arisen in this way) produces some of the symptoms of schizophrenia, perhaps rigidity and repetition. I am not claiming that schizophrenia is simply consciously controlled. It is in the beginning, but later the process is irreversible by the patient's efforts. For one thing, he soon obscures to his own view the path by which he was led to the disease. His libido soon invests itself in perversions and the struggle against them. He loses interest in succeeding in the normal social world. He in time represses value judgments which are incompatible with "hiding out," within the confines of his own mind, from society.

His primary mental playing-dead mechanism is partially automatic and never was in his full control, and has been supplemented by a kind of secondary neurosis. He cannot will himself to become well. But he *can* tell another person what he is afraid of, and this is the first step to cure.

A considerable part of schizophrenic queerness in thought and behavior is "phony"—i.e., consciously, strategically adopted to mislead observers in the interest of concealing the all-important truth. Another part of it is "genuine," unconsciously originated or consciously originated but sincerely believed in. I suggest that when a large part of a person's libidinal and nonlibidinal drives are repressed, the remaining ego is so slim and weak as to easily accept falsehood for truth, and any falsehood or false perception which relieves tension is apt to be believed. But it should be emphasized that the schizophrenic is consciously uncooperative and insincere—is apt to be "as slick as an eel in a bucket of oysters"—and therapy must take this into account.

Additional important points neglected in the above:

a. Deliberate insincerity equals "split personality."

b. Delusions are but weakly believed by the patient himself when alone or unthreatened. Delusions are dormant except in interpersonal relationships, where they are trotted out and strongly advanced as weapons of defense.

c. The more punitive, threatening, and coercive the society (or segment of society), the greater the withdrawal from intimacy (in degree and range).

d. The lower the stake in normal sociality (promise of future happiness) judged largely from past experience, the greater the withdrawal from intimacy.

e. Withdrawal from intimacy is self-accelerating, since partial withdrawal lowers the stake in normal sociality.

f. The greater the acquired skill in enjoying solitary pleasures (say, reading), the greater the withdrawal from intimacy.

g. The more sudden and more contrasted the change from a less to more punitive segment of society, the greater the withdrawal from intimacy.

h. The more cohesive the society with regard to "informing against" wrongdoers, the greater the withdrawal from intimacy.

i. The greater the known number of secret policemen, detectives, informers, spies, etc. the greater the withdrawal from intimacy. Thus, a modern totalitarian state and wartime armies should produce a large percentage of schizophrenics.

Schizophrenic withdrawal is *not* merely from authoritative persons, say, parents, policemen, or doctors. It tends to be a withdrawal of intimacy from *everybody*. The rich interrelatedness and intercommunication of modern society and centralization of authority (as contrasted with former historical periods) tend to produce schizophrenia.

My theory makes schizophrenics out, in a sense, to be villains. Yet obviously they would be bungling villains, and the very botching of their lives strongly suggests forceful extenuating circumstances, since without tremendous pressures people seldom, presumptively, destroy what they love as much as all people love themselves originally.

Irrelevantly, it is pleasant to say that human beings are lent dignity (of a rather evil sort, if my theory is true) if one credits the schizophrenics of the human race with being real people who have fallen into disease by having failed to make the right decisions at the right time in regard to problems which even our philosophers and theologians have not solved in the quiet of their studies with the help of the written words of many other minds—while the schizophrenics failed under enormous, frantic pressures, as is typically the case.

The gradual transition in degrees of mental sickness from the worst to the best cases in a mental hospital and on to the most sick and then least sick of normal people, argues against mental illness having simple physical causes. The people with pneumonia in a community make no symptomatic transition whatever with those not suffering from pneumonia.

Military life increases dammed up sexuality (especially homosexuality) and hostility (the natural product of fear and coercion). I offer that the greater the degree of preceding neurosis, the greater the facility with which the subject develops schizophrenia, although libidinal repression is not the primary cause of the disease.

Why the high incidence of schizophrenia in the armed forces? If my Dick Tracy theory is true, several factors explain this. Soldiers are deprived of privacy, they live on view like goldfish in a bowl. In time they can virtually read one another's minds. Thus unusual measures of insincerity are called for to keep important secrets concealed. And the opportunities for wrongdoing increase tremendously, especially in the pressured confusion of wartime (such as theft of government property, rape and murder of civilians in occupied countries, semiaccidental injury of one's fellow soldiers in the heat of combat, etc.).

Also in wartime, it becomes customary to suspend normal scruples, which are apt later to rush back with a vengeance. At the same time, the armed forces constantly "advertise" the severity of their punishments, increasing fear of wrongdoing and detection—and indirectly, increasing insincerity in the soldiers, among whom wrongdoing abounds. Because forced intimacy is one factor in schizophrenia, I think the present practice of throwing patients together in mental hospitals is not ideally therapeutic (though it may be an economic necessity).

It is doubtful whether merely diagnostic skill or the incidence of schizophrenia is increasing with the decades. I suspect the disease itself is increasing, and the increase can be explained thus: in the increasing complexity and pressure in society, wrongdoing is itself increasing. At the same time, due to closed frontiers, improved communications, etc. police power and ability to detect wrongdoing are increasing. Thus fear of detection is an increasingly common experience.

THOUGHTS ON THERAPY

Since in the four types of schizophrenia the causative virus is a more or less conscious fear of detection and punishment, the way is pointed to successful therapy. A guarantee of amnesty for the misdeeds and the exposure of the misdeeds would take away entirely the core motive for the illness, and a rather rapid retrogression of the libido to genitality should take place.[3]

Among the patients of a large mental hospital, it seems likely that some of the schizophrenics are concealing crimes as serious as first-degree murder. If society, through its legislatures and judges, is not willing to grant amnesty to all schizophrenics, then (assuming, of course, that my theory is correct) the outlook for therapy might remain as dismal as it is at present. On the other hand, a procedure of exposing and punishing misdeeds might be somewhat therapeutic and even more humane, in the long run, than the present procedure of waiting for the patients to "snap out of it" or applying therapy which does not seem to do much permanent good for the schizophrenic. I suppose many people wait hopefully for some far-off scientist to discover the physical aspects of primarily attitudinal and ideational disorders, and the drugs and other physical means which will cure them.

My belief is that both in origin and in respect to thorough cure, functional psychoses are purely mental to a large extent which will remain even after physical help is found. The patient's long, unnatural efforts at stifling sexual drive, emotion, curiosity, activity, etc., all as being dangerous to self-preservation or safety, and the long-contained state of terror in which they live (pickling the nerves in adrenalin, as it were) are likely to cause some physiological malfunctioning. And undoubtedly there is some physical basis of the "weakness" which many of these poor people evidence in committing antisocial deeds in the first place, and in hastening to the employment of a last-ditch defense mechanism to hide deeds which stronger people might have laughed off. If drugs hasten the realization in the schizophrenic that danger has passed or else never existed, or else that the danger does exist and is important but that its consequences are preferable alternatives to their present agonized state of dismal prospects—for many of them are in a state probably "worse than death"—if drugs help these realizations, then well and good. They can be used as adjuncts. The purely mental approach may remain an indispensable tool.

I have a suggestion for more effective therapy for schizophrenics. A *group* of therapists working with one patient at a time, even for brief sessions, could find the heart of the trouble rapidly. The only danger is that the patient might be driven deeper into his defenses—but the therapists would stop the treatment for a time in that case. The therapists should do much of the talking. Such a procedure brings to mind—judge and jury, the family circle, the classroom, the third degree, and brainwashing techniques! But it might work.

This method might be called surgical psychotherapy, since the aim would be rapid excision of root troubles.

A group has the unique power of supplanting for a time the individual's superego—and it can generate new emotional values in the patient on the spot.

I believe perhaps thousands of psychotherapists and relatives of patients, etc. around the world have adopted a kind of Dick Tracy theory as a working hypothesis in dealing with at least some schizophrenics—but to my knowledge no one has developed the theory (enlarged and fitted its details into the data of authoritative knowledge).

Perhaps "brainwashing" techniques would be therapeutic for schizophrenia, since there is a need for conversion from an *antisocial philosophy*. As ministers of religion know, reform and enduring conversion are not easily obtained.

Since psychotherapy is easily viewed by the patient as threatening exposure, insincerity and fear often increase in the schizophrenic in psychotherapy, accounting for the fact that psychotherapy in this case often has an effect opposite to that intended; namely, increase in severity of the disease.

Perhaps some new liaison between psychotherapists and, on the other hand, law enforcement agencies, lawyers, criminologists, ministers—might be helpful.

A direct condemnatory approach on the part of the psychotherapist might, indeed, tend to make the patient worse. So I'm not sure what approach should be used.

It is the firmly entrenched antisocial (self-protective) ideology behind schizophrenia (philosophically and religiously supported insincerity) which makes permanent, real cure in schizophrenia so difficult, makes the disease so stubborn.

I have an idea for an adjunctive type of therapy for schizophrenia. It misses the essential problem—concealment—but helps in rebuilding a lost ego, or one seriously narrowed. My idea is for educational bibliotherapy of a special kind, which may be called encyclopedic therapy. Since, typically, many sane years were spent in school, a kind of review of all knowledge helps to recover healthy memories and associations. Therapy devoted to one field of knowledge does not do this. Encyclopedias and dictionaries afford a complete survey of reality. There exist at present popular small dictionaries and brief, one-volume encyclopedias.

---

[3] This emphasis upon the importance of "amnesty" may be quite unrealistic, as other comments of the author himself indicate. Already he has remarked upon the lack of established means in our society for the controlled resolution of guilt—through restitution, sacrifice, or the like. *Self*-condemnation is probably a more immediate problem here than is social judgment; and it is by no means certain that formal social forgiveness necessarily implies self-forgiveness. See also later parts of this section.—O. H. M.

It should be well noted that at the present time most patients in mental hospitals are incarcerated and are "serving time" *without* in any way earning credit toward ultimate forgiveness by society.

It should be noted that even after the removal of the core motive for the disease, the therapeutic effect would not show up for a long time, because the patient is thoroughly habituated to his narrowed psychic and social life and has actually forgotten many of the normal adjustments to life.

Insincerity with regard to society is an *alternative* to repression. But repression may be regarded as a self-sincerity.

Insincerity stands out as much in schizophrenia as does repression in neurosis.

I suspect the steady diet of TV programs emphasizing crime, detection, and punishment is usually detrimental to the schizophrenics who absorb it several hours a day currently in mental hospitals.

I believe that society itself has a kind of split personality with regard to punishment for wrongdoing. On the one hand, Christ and Christianity are extolled; on the other hand, policemen and judges are venerated. Christ's enjoinders, judge not, forgive, turn the other cheek, love! are seriously at odds with the judgment and punishment of criminals. I believe few people are aware of this conflict—but it is there, nevertheless, as an academic problem for innocent people, and as a serious emotional problem for scrupulous undetected wrongdoers—the very class which falls prey to schizophrenia. The exact meaning of the Gospels is not clear with regard to specific action. But this only makes the conflict worse. Most people are *not* master theologians. The United States flourished for more than 50 years with a similar split in its personality—with a conflict between its Declaration of Independence, Constitution, Bill of Rights on one hand, and degraded slavery of Negroes on the other. I believe the conflict with regard to concealing and forgiving or revealing and punishing is, although given little recognition—and being invisible or nonexistent to many people—a most serious source of emotional maladies. Many believe that conflict is central in mental disease. If it is, this conflict is the greatest of all in schizophrenia.

Since neither the institution of Christianity nor the institution of the police gives evidence of going out of business, perhaps the solution to this social problem lies in giving emphasis to a clean-cut assertion of the relevance of Christianity only to sins not punishable by law. But on what philosophical and theological grounds this assertion would be made, it is doubtful. Thus schizophrenia is in part a sociological problem.

So what does the schizophrenic's underlying nature point to in the way of therapy?

Primarily, fear must be reduced. The tranquilizing drugs help (temporarily) in this.

A number of secondary points show up. Reduction of guilt feelings. Increase of interest in other people and need for companionship. Increase in honesty. Increase in awareness of being sick and of living a life not worth living. The therapist could guarantee due process of law, legal counsel, etc. Standard type psychotherapy could be used later. Whereas therapy of neurosis requires that the patient do most of the talking, at the beginning the therapist should do most of the talking.

Detectives and priests and lawyers and doctors all know that the need to confess important secrets—even in the face of danger—is very great. The need to confess is not shared by men with animals. This specifically human need to become, through communication, honestly reconciled to the "offended" group is, I conjecture, in effect tantamount to a primary impulse—like sexual drive. The unconfessed wrongdoer is self-excommunicated from society—and his loneliness is great. This is a major pain of criminals and schizophrenics. And I conjecture that the need of confession has a specific goal; namely, those people most concerned with the wrongdoing. Thus confession, say, of a crime to a doctor's ears but not to police, is apt to do little good, the need not being satisfiable by those not primarily concerned.

A suppressed or repressed secret tends to grow like a tumor. "Secrets" behave just like dammed up emotions or impulses in neurosis. A secret misdemeanor becomes, in the recesses of the mind, tantamount to a felony.

No wonder mental hygienists emphasize the need of confidantes before whom there are no important secrets.

And morality may get a boost some day through science. It may be discovered that wrongdoing, in addition to its other faults, is severely unhealthy—and is to be avoided "like the plague."

If the Dick Tracy theory is true, if a schizophrenic were to lose all inhibitions, observers would witness torrential outbursts of fear and hatred of punishment. Of secondary intensity, an outburst of confession—of guilt, need for punishment, need for forgiveness and of least intensity but still intense, an outburst of perverted sexual interest—homosexuality, sado-masochism, and urethral and oral and exhibitionistic drives.

Perhaps the pillory and stocks placed in public are the best means of punishment, the most healthy and effective and efficient—and society may, through science, return to using them.

SUMMARY

In a nutshell, my theory disagrees, so far as I know, with previously developed and circulated theories of schizophrenia, in that it rejects the concept of the primary importance of repression—I know Rosanoff said schizophrenics have too few repressions, but as far as I know, he made no systematic development of the principle. Rather, my idea is that repressed impulses—sexual and otherwise libidinal, and hostile—occur later, secondarily, after a mental play-dead mechanism is triggered by the need to hide likely punishable wrongdoing from other people. The pressure of being unshriven with regard to matters of the utmost emotional importance to the patient is great. And since mental associations naturally tend to ramify, it is no easy task to live in close intimacy and keep our intimates from guessing about the cause of our anxiety. The schizophrenic throws overboard the very appearance of anxiety, he throws overboard his friendships, retreats from intimacy itself—this mechanism is so cunning, is so readily called into play by even uneducated people living in different cultures in different parts of the world, that it must be an instinctive mechanism. No one needs to learn, any more than his body needs to learn the state of shock, how to fool anybody and everybody, using supreme animal cunning. If my theory is true, schizophrenics around the world throughout centuries have succeeded superbly well in dissimulating wrongdoing, for even now in 1956, even psychiatrists have not yet learned about schizophrenics, "what they are up to."

My theory, if not true, is at least hauntingly self-consistent and is not definitely at odds with any strongly supported conceptions, so far as I know. Perhaps my theory is erroneous but still might fertilize some other minds to solve the problem of schizophrenia. At worst, my theory is a cunningly contrived, systematic, self-consistent set of errors in some harmony with authoritative knowledge.

My Dick Tracy theory may be thought of as a combining of the common view held early in World War I that military psychotics were feigning sickness in order to escape duty, with the present common view that schizophrenia is a disease like cancer, for which the patient is no way morally responsible, the disease having an as yet to be discovered organic cause. However, my theory did not originate in this fashion. In the matter of truths, all roads lead to Rome. Twelve years of ruthless self-analysis led to my theory.

My view is that the schizophrenic has adopted semi-instinctive, insincere stratagems in order to escape punishment, one of them being withdrawal from intimacy—which isolation causes a libidinal starvation resulting in an involuntary, neurotic-type outbreak of symptoms based on unconscious, pregenital (perverted libidinal) drives. And many military cases of schizophrenia are overdetermined by feigning in order to escape duty.

So what is a schizophrenic? In brief, he is a terrified, conscience-stricken crook, who has repressed his interest in people, unavowedly insincere and uncooperative, struggling against unconscious sexual perversion. He is of no mean Thespian ability. And his favorite Commandment is that which one nowadays facetiously calls the Eleventh Commandment, "Thou shalt not get caught."

Attempts to expose him may only drive him further "underground." But a knowledge of his true nature will surely lead, someday, to someone's discovering a sure, quick, effective, and enduring cure.

* * * * *

My "qualifications" for developing my theory:

Nearly four years of psychotherapy.

Nearly two years of living in mental hospitals.

Observation of my mother, who is now hospitalized for the fourth time.

Twelve to fifteen years of self-analysis.

A B.A. degree and one year of graduate work in psychology with philosophy as a minor.

99th percentile on the Graduate Record Examination in psychology.

A nearly lifelong desire to contribute to knowledge, especially to solve the problem of mental illness.

Received November 14, 1957.

CRITIQUE AND NOTES

VOCABULARY RESPONSE AS RELATED TO LATERALIZATION OF CEREBRAL LESION: AN INVESTIGATION OF "LATENT APHASIA"

ALFRED B. HEILBRUN, JR.
State University of Iowa

In a recent study, Pichot (7) compared the performance of senile dements and arteriosclerotic patients on a Wechsler-like vocabulary test (6) and the Progressive Matrices test (6). He found, following a correction for age, no difference in performance on the test of spatial reasoning ability, whereas the arteriosclerotic group scored significantly lower on the vocabulary measure. Since all *S*s who demonstrated clinically detectable dysphasic symptomatology had been eliminated from both groups, he felt that this pattern of scores suggested the presence of a dysphasic condition of sufficient intensity to be evidenced on the tests but not upon neurologic examination. Pichot suggested the term "latent aphasia" to describe this condition, pointing out how it might be fruitfully used to account for the impaired verbal test performances of epileptic patients reported by previous workers (1, 9).

Heilbrun (4) found that after elimination of patients showing clinically evident dysphasia, right-handed *S*s judged to have cerebral pathology primarily involving the left hemisphere displayed a significantly different pattern of performance on "verbal" and "spatial" batteries than did a comparable group with right hemisphere involvement. The left-sided *S*s did more poorly on the "verbal" battery as compared with "spatial," whereas the right-sided *S*s showed the opposite pattern. Since there is a highly stable relationship between dysphasia and left cerebral pathology in right-handed persons, this might be considered evidence for the presence of a subclinical dysphasic condition in the left-sided group. However, there was no significant difference between groups on the Wechsler-Bellevue Vocabulary test, which was part of the "verbal" battery.

McFie and Piercy (5) reported that after excluding dysphasic *S*s from consideration, the left-sided *S*s scored significantly lower on a verbal task (Wechsler-Bellevue Similarities Test) than did right-sided *S*s.

The present study was conducted to evaluate the concept of "latent aphasia" by comparing the vocabulary test performance of *S*s judged to have predominantly left-sided cerebral pathology, but without clinically evident dysphasia, with that of *S*s judged to have pathology primarily involving the right hemisphere, and with a control group of hospitalized *S*s with no suspected brain pathology. Since an earlier analysis of vocabulary performance (4) had demonstrated no difference between grossly nondysphasic left-sided *S*s and right-sided *S*s when standard scoring procedures were used, this study investigated the ability of a special judgmentally derived scoring system to reflect differences in these groups.

METHOD

Subjects. The 44 *S*s included in this study were all patients in the University Hospital, State University of Iowa, or in the Veterans Administration Hospital in Iowa City. They were right-handed, nonpsychotic, American-born adults of both sexes.

The *S*s were categorized into three groups based on the expressed judgment of a staff neurologist or neurosurgeon as to primary lateral locus of cerebral lesion (LS group, $n = 13$; RS group, $n = 14$) or the opinion of the physician on a non-neurological ward that there was no indication of cerebral pathology (C group, $n = 17$). There were no significant differences between the LS, RS, and C groups with respect to mean age (49, 48, and 47, respectively) or mean verbal IQ as measured by the Wechsler-Bellevue with Vocabulary omitted (94, 99, and 98, respectively).

Vocabulary Test. The verbal task used in this study was the Vocabulary subtest of the Wechsler-Bellevue, Form I. However, performance was not evaluated by the usual right-wrong scoring method but in terms of the qualitative mode of response utilized by *S* independent of the accuracy of his answer. It was felt that an analysis by mode of response might be more sensitive to subtle decrements in verbal proficiency than by the usual scoring system. Other studies (2, 3) have found that when more qualitative measures of vocabulary performance are used, neuropsychiatric groups tend to be inferior to normal controls on this type of task.

Fourteen Ph.D. clinical psychologists were asked to rank the following four modes of vocabulary responses along a continuum of intellectual adequacy: (*a*) categorical ("A donkey is an animal"), (*b*) descriptive ("A donkey has long ears and four legs"), (*c*) equivalent ("A donkey is a jackass"), and (*d*) functional ("You ride on a donkey"). Mean rankings defined three levels of adequacy: Lever I (high), categorical; Level II, equivalent; Level III (low), descriptive or functional. There was 86 per cent agreement as to these relative positions among the 14 raters.

The 44 completed vocabulary tests were then presented individually to six advanced graduate students in clinical psychology, and each was asked to rate any attempted noun item (whether correct or incorrect) as to mode of response. There are 28 such items on the test. No information was given as to the nature of the

TABLE 1
COMPARISON OF MEAN ITEM AND MEAN TOTAL SCORES OF LEFT-SIDED (LS), RIGHT-SIDED (RS), AND CONTROL (C) GROUPS ON WECHSLER-BELLEVUE VOCABULARY

Group	Test Scored by Mode of Response			
	Mean Vocabulary Item Score	Standard Deviation	Mean Total Vocabulary Score	Standard Deviation
LS	2.3	.46	49.5	12.3
RS	2.3	.71	40.3	17.7
C	2.3	.44	43.2	11.5

*S*s or the intent of the research. In cases where an item seemed to include more than one mode of response, the judges were instructed to make multiple mode ratings. In such cases, the mode which fell into the highest level was arbitrarily chosen as characterizing the response. Interjudge agreement for level of response was approximately 84 per cent.

The mean rating for the six judges was used to determine the final level assigned to each vocabulary response. For statistical evaluation, each response assigned to Level I was given a weight of 3; to Level II, a weight of 2, and to Level III, a weight of 1.

RESULTS

The mean item and mean total scores for the three groups are presented in Table 1. None of the comparisons of these group means produced a difference significant at the 10 per cent level of confidence.

These results fail to support the notion of "latent aphasia." The group with judged left hemisphere pathology but no clinically observable dysphasia did not perform less adequately than either the right-sided or the control group.

DISCUSSION

The failure to find evidence of "latent aphasia," in contrast to Pichot's report, may be attributed to several differences in the two studies. For one, Pichot used *S*s whose medical classifications imply diffuse deterioration of the cerebral cortex. However, pathology of the left cerebral hemisphere, a prerequisite for "latent aphasia," was an inference based on criterion task performance. In the present study, lateralization of lesion was based on expert judgment, independent of vocabulary performance.

A second factor of possible importance was the marked age differences between Pichot's *S*s (mean age arteriosclerotic group = 65 years; mean age senile group = 73 years) and those utilized in this study. There may be characteristics (e.g., motivational status, ability to concentrate, etc.) associated with arteriosclerotic disease in the aged which could account for the obtained differences from senile *S*s on a verbally presented vocabulary task and the finding of no difference on a performance measure where the task stimuli are continuously present.

A third difference is in the scoring technique for the criterion measure. The present study used a more qualitative evaluation of the vocabulary responses while the standard right-wrong system was utilized by Pichot.

The findings of the present study cast some doubt on the utility of the concept of "latent aphasia." If psychological testing is to prove of some value in detecting and measuring minimal dysphasic conditions, it would seem more promising to approach the problem by attempting to construct verbal tests specifically for this purpose rather than to use present tests which were originally devised for other uses (e.g., tests of intelligence). An empirical approach might be utilized in which a large pool of verbal items would be administered to *S*s judged to show minimal dysphasia and *S*s judged to be nondysphasic. Those items which demonstrate an ability to discriminate could then be included in a further examination of "latent aphasia."

SUMMARY

The performances of two brain-injured groups and a control group were compared on the Vocabulary subtest of the Wechsler-Bellevue, Form I. One brain-injured group was judged to have cerebral pathology primarily involving the left hemisphere and the other primarily involving the right. The control group was judged to be free of cerebral pathology. Further, the left-sided group did not demonstrate manifest dysphasic symptoms.

The noun items on the Vocabulary test were scored by mode of response (categorical, descriptive, equivalent, or functional) independent of accuracy. Analyses by both item scores and mean total scores demonstrated no differences between these groups.

These results failed to yield evidence suggestive of Pichot's concept of "latent aphasia" as defined by impaired performance on a Wechsler-like vocabulary test in the absence of clinically evident dysphasia. It was suggested that further research might be more fruitful if psychological tests were specifically constructed to measure minimal dysphasia by empirical techniques rather than continuing to use present tests devised for other purposes.

REFERENCES

1. CAPPS, H. M. Vocabulary changes in mental deterioration. *Arch. Psychol.*, 1939, No. 242.
2. CHODORKOFF, B., & MUSSEN, P. Qualitative aspects of the vocabulary responses of normals and

schizophrenics. *J. consult. Psychol.*, 1952, **16**, 43–48.

3. FEIFEL, H. Qualitative differences in the vocabulary responses of normals and abnormals. *Genet. psychol. Monogr.*, 1949, **39**, 151–204.
4. HEILBRUN, A. Psychological test performance as a function of lateral localization of cerebral lesion. *J. comp. physiol. Psychol.*, 1956, **49**, 10–14.
5. MCFIE, J., & PIERCY, M. F. Intellectual impairment with localized cerebral lesions. *Brain*, 1952, **75**, 292–311.
6. PENROSE, L. S., & RAVEN, J. C. A new series of perceptual tests. *Brit. J. med. Psychol.*, 1936, **16**, 97–104.
7. PICHOT, P. Language disturbances in cerebral disease. *Arch. Neurol. & Psychiat.*, 1955, **74**, 1–5.

Received June 17, 1957.

RESPONSE SELECTIONS OF NEUROTIC AND CONTROL SUBJECTS AS A FUNCTION OF RELATIVE FAMILIARITY AND SOCIAL DESIRABILITY OF ITEMS

H. J. WAHLER[1]

Veterans Administration Hospital, Knoxville, Iowa

A NUMBER of recent studies (3) have shown that positively valued stimuli tend to facilitate or "sensitize" perceptual recognition. Conversely, Bruner and Postman (1) and others (4) have reported findings supporting the proposition that negatively valued, affectively charged, or taboo stimulus words produce an inhibitory effect on recognition. There are also related studies (7, 11) providing evidence that *S*s tend to inhibit or withhold socially disapproved responses.

Various principles have been hypothesized to account for the experimental findings, but none of them has attained unequivocal status. Soloman and Howes (9) attempted to account for both the effects of positive and negative valued stimuli in terms of a word-frequency hypothesis. Postman recognized the significance of frequency as an important factor in the observed selectivity but also contended that experiments in which frequency effects were controlled produce "... evidence that the frequency cannot by any means account for the total of all the observed facts of perceptual selectivity ..." (6, p. 81). In a recent review of studies in this area, Jenkins (3) argued for the contribution of other factors than frequency. Thus, both frequency or, more generally, familiarity and other associated properties of stimuli may be important determinations of perceptual recognition thresholds and response selections or preferences.

Employing a scrambled sentences task, which in effect constituted a disguised choice situation with neutral and hostile alternatives, Watson, Pritzker, and Madison (10) studied the response selections of neurotic and control groups to determine whether neurotic *S*s preferred hostile to nonhostile alternatives to a greater extent than controls. In addition to demonstrating this relationship, the authors also noted that some of their items elicited many more hostile selections from *both* groups than did others. They hypothesized that these inter-item effects were attributable to the relative familiarity of the hostile and neutral alternatives and, in this connection, cited the findings of Soloman and Howes (9) and Postman and Crutchfield (8) to substantiate their view.

In light of the evidence supporting the position that both positive and negative values associated with stimuli serve to facilitate or inhibit response selection in addition to the effects of frequency, it is important to study the independent contribution of both variables to the variance in response selection. Furthermore, it is of importance to personality theory to determine whether the possible perceptual selectivity effects of social values operate differentially with neurotic and control groups. Mowrer (5) has proposed that neurotics have learned the values and attitudes of society but are unable or unwilling to apply them. This suggests the hypothesis that the response preferences of neurotics are less strongly related to social values than are those of nonneurotic people.

The purpose of this note is to examine the contribution of social desirability and familiarity to the variance in response preferences to a choice task when the effects of each is controlled. An additional purpose is to compare the effects of relative social desirability and familiarity in relation to the response preferences of neurotic and control groups.

METHOD

The 120 possible alternative sentences from the list presented by Watson, Pritzker, and Madison (10) were rated for familiarity on a nine-point scale. Ebel's (2) index of rater reliability ranged between .93 and .98 for three groups of raters ($N = 23$, 9, and 14). Correlations

---

[1] The author wishes to express his appreciation to A. L. Benton, H. P. Bechtoldt, L. D. Goodstein, and F. E. Ash for their many helpful suggestions.

TABLE 1

COMPARISON OF MEAN ITEM AND MEAN TOTAL SCORES OF LEFT-SIDED (LS), RIGHT-SIDED (RS), AND CONTROL (C) GROUPS ON WECHSLER-BELLEVUE VOCABULARY

Group	Test Scored by Mode of Response				
	Mean Vocabulary Item Score	Standard Deviation	Mean Total Vocabulary Score	Standard Deviation
LS	2.3	.46	49.5	12.3
RS	2.3	.71	40.3	17.7
C	2.3	.44	43.2	11.5

*S*s or the intent of the research. In cases where an item seemed to include more than one mode of response, the judges were instructed to make multiple mode ratings. In such cases, the mode which fell into the highest level was arbitrarily chosen as characterizing the response. Interjudge agreement for level of response was approximately 84 per cent.

The mean rating for the six judges was used to determine the final level assigned to each vocabulary response. For statistical evaluation, each response assigned to Level I was given a weight of 3; to Level II, a weight of 2, and to Level III, a weight of 1.

RESULTS

The mean item and mean total scores for the three groups are presented in Table 1. None of the comparisons of these group means produced a difference significant at the 10 per cent level of confidence.

These results fail to support the notion of "latent aphasia." The group with judged left hemisphere pathology but no clinically observable dysphasia did not perform less adequately than either the right-sided or the control group.

DISCUSSION

The failure to find evidence of "latent aphasia," in contrast to Pichot's report, may be attributed to several differences in the two studies. For one, Pichot used *S*s whose medical classifications imply diffuse deterioration of the cerebral cortex. However, pathology of the left cerebral hemisphere, a prerequisite for "latent aphasia," was an inference based on criterion task performance. In the present study, lateralization of lesion was based on expert judgment, independent of vocabulary performance.

A second factor of possible importance was the marked age differences between Pichot's *S*s (mean age arteriosclerotic group = 65 years; mean age senile group = 73 years) and those utilized in this study. There may be characteristics (e.g., motivational status, ability to concentrate, etc.) associated with arteriosclerotic disease in the aged which could account for the obtained differences from senile *S*s on a verbally presented vocabulary task and the finding of no difference on a performance measure where the task stimuli are continuously present.

A third difference is in the scoring technique for the criterion measure. The present study used a more qualitative evaluation of the vocabulary responses while the standard right-wrong system was utilized by Pichot.

The findings of the present study cast some doubt on the utility of the concept of "latent aphasia." If psychological testing is to prove of some value in detecting and measuring minimal dysphasic conditions, it would seem more promising to approach the problem by attempting to construct verbal tests specifically for this purpose rather than to use present tests which were originally devised for other uses (e.g., tests of intelligence). An empirical approach might be utilized in which a large pool of verbal items would be administered to *S*s judged to show minimal dysphasia and *S*s judged to be nondysphasic. Those items which demonstrate an ability to discriminate could then be included in a further examination of "latent aphasia."

SUMMARY

The performances of two brain-injured groups and a control group were compared on the Vocabulary subtest of the Wechsler-Bellevue, Form I. One brain-injured group was judged to have cerebral pathology primarily involving the left hemisphere and the other primarily involving the right. The control group was judged to be free of cerebral pathology. Further, the left-sided group did not demonstrate manifest dysphasic symptoms.

The noun items on the Vocabulary test were scored by mode of response (categorical, descriptive, equivalent, or functional) independent of accuracy. Analyses by both item scores and mean total scores demonstrated no differences between these groups.

These results failed to yield evidence suggestive of Pichot's concept of "latent aphasia" as defined by impaired performance on a Wechsler-like vocabulary test in the absence of clinically evident dysphasia. It was suggested that further research might be more fruitful if psychological tests were specifically constructed to measure minimal dysphasia by empirical techniques rather than continuing to use present tests devised for other purposes.

REFERENCES

1. CAPPS, H. M. Vocabulary changes in mental deterioration. *Arch. Psychol.*, 1939, No. 242.
2. CHODORKOFF, B., & MUSSEN, P. Qualitative aspects of the vocabulary responses of normals and

schizophrenics. *J. consult. Psychol.*, 1952, **16**, 43–48.
3. FEIFEL, H. Qualitative differences in the vocabulary responses of normals and abnormals. *Genet. psychol. Monogr.*, 1949, **39**, 151–204.
4. HEILBRUN, A. Psychological test performance as a function of lateral localization of cerebral lesion. *J. comp. physiol. Psychol.*, 1956, **49**, 10–14.
5. MCFIE, J., & PIERCY, M. F. Intellectual impairment with localized cerebral lesions. *Brain*, 1952, **75**, 292–311.
6. PENROSE, L. S., & RAVEN, J. C. A new series of perceptual tests. *Brit. J. med. Psychol.*, 1936, **16**, 97–104.
7. PICHOT, P. Language disturbances in cerebral disease. *Arch. Neurol. & Psychiat.*, 1955, **74**, 1–5.

Received June 17, 1957.

RESPONSE SELECTIONS OF NEUROTIC AND CONTROL SUBJECTS AS A FUNCTION OF RELATIVE FAMILIARITY AND SOCIAL DESIRABILITY OF ITEMS

H. J. WAHLER[1]

Veterans Administration Hospital, Knoxville, Iowa

A NUMBER of recent studies (3) have shown that positively valued stimuli tend to facilitate or "sensitize" perceptual recognition. Conversely, Bruner and Postman (1) and others (4) have reported findings supporting the proposition that negatively valued, affectively charged, or taboo stimulus words produce an inhibitory effect on recognition. There are also related studies (7, 11) providing evidence that *S*s tend to inhibit or withhold socially disapproved responses.

Various principles have been hypothesized to account for the experimental findings, but none of them has attained unequivocal status. Soloman and Howes (9) attempted to account for both the effects of positive and negative valued stimuli in terms of a word-frequency hypothesis. Postman recognized the significance of frequency as an important factor in the observed selectivity but also contended that experiments in which frequency effects were controlled produce "... evidence that the frequency cannot by any means account for the total of all the observed facts of perceptual selectivity ..." (6, p. 81). In a recent review of studies in this area, Jenkins (3) argued for the contribution of other factors than frequency. Thus, both frequency or, more generally, familiarity and other associated properties of stimuli may be important determinations of perceptual recognition thresholds and response selections or preferences.

Employing a scrambled sentences task, which in effect constituted a disguised choice situation with neutral and hostile alternatives, Watson, Pritzker, and Madison (10) studied the response selections of neurotic and control groups to determine whether neurotic *S*s preferred hostile to nonhostile alternatives to a greater extent than controls. In addition to demonstrating this relationship, the authors also noted that some of their items elicited many more hostile selections from *both* groups than did others. They hypothesized that these inter-item effects were attributable to the relative familiarity of the hostile and neutral alternatives and, in this connection, cited the findings of Soloman and Howes (9) and Postman and Crutchfield (8) to substantiate their view.

In light of the evidence supporting the position that both positive and negative values associated with stimuli serve to facilitate or inhibit response selection in addition to the effects of frequency, it is important to study the independent contribution of both variables to the variance in response selection. Furthermore, it is of importance to personality theory to determine whether the possible perceptual selectivity effects of social values operate differentially with neurotic and control groups. Mowrer (5) has proposed that neurotics have learned the values and attitudes of society but are unable or unwilling to apply them. This suggests the hypothesis that the response preferences of neurotics are less strongly related to social values than are those of nonneurotic people.

The purpose of this note is to examine the contribution of social desirability and familiarity to the variance in response preferences to a choice task when the effects of each is controlled. An additional purpose is to compare the effects of relative social desirability and familiarity in relation to the response preferences of neurotic and control groups.

METHOD

The 120 possible alternative sentences from the list presented by Watson, Pritzker, and Madison (10) were rated for familiarity on a nine-point scale. Ebel's (2) index of rater reliability ranged between .93 and .98 for three groups of raters (N = 23, 9, and 14). Correlations

---

[1] The author wishes to express his appreciation to A. L. Benton, H. P. Bechtoldt, L. D. Goodstein, and F. E. Ash for their many helpful suggestions.

TABLE 1

INTERCORRELATIONS AMONG RELATIVE FAMILIARITY, SOCIAL DESIRABILITY, AND RESPONSE PREFERENCES OF NEUROTIC AND CONTROL GROUPS

	Social desirability	Frequency of hostile selections: Neurotics	Frequency of hostile selections: Controls
Familiarity	.708*	−.610	−.656
Social desirability		−.496	−.644

* *r*s of .330 or higher are significant at the .01 level of confidence; $N = 60$ for all coefficients.

of the average ratings of all combinations of the three groups ranged between .90 and .91. These indices indicate very high intra- and intergroup agreement. The average pooled ratings of the three groups were used as indices of familiarity for each alternative sentence. The social desirability ratings of the same items were obtained from a fourth group ($N = 23$). The index of rater reliability for these ratings was .92.

To obtain relative values of the two alternatives on each item for rated familiarity and rated social desirability, difference scores for all items were computed. These scores were obtained for both variables by subtracting the relevant value of the hostile alternative from the nonhostile alternative for each item. The reported frequencies with which hostile alternatives were selected by the two groups were used as indices of strength of preference.

RESULTS AND DISCUSSION

From Table 1 it may be seen that the intercorrelations among the variables of relative familiarity, social desirability, and frequency of hostile selections by neurotic and control *S*s are all significant beyond the .01 level. None of the negative correlations differ significantly from each other.

To determine whether both familiarity and social desirability contribute significantly to the variance in preferences for alternatives, partial correlations were computed for both groups of *S*s. With nonneurotic *S*s, when familiarity of alternatives was thus controlled statistically, the relationship between relative social desirability and frequency of choices remained significant ($r = -.338$; $p < .01$). The partial correlation between familiarity and frequency of choices with social desirability controlled was −.370. For neurotic *S*s, when the variance attributable to familiarity was partialled out, the relationship between relative social desirability and frequency of hostile choices was no longer significant ($r = -.115$). With relative social desirability partialled out the relationship between familiarity and response preferences remained significant ($r = -.422$; $p < .01$). The *differences* among all combinations of these four partial *r*'s, however, were not sufficiently large to attain statistical significance.

These findings indicate that, with a choice task of the type used by Watson *et al.*, both the relative familiarity and social desirability of alternatives contribute significantly to the variance in the response selection of control *S*s. The findings also show that both of these variables may function as general response determinants in a choice task.

The contrasting lack of evidence that response preferences of neurotic *S*s are related to social desirability when familiarity is controlled suggests, as proposed by Mowrer, that neurotic and nonneurotic people may tend to respond differently to social values. It will be recalled, however, that none of the partial correlation coefficients given above differed significantly. Thus, this proposition is only supported by the direction of the results.

SUMMARY

The primary purpose of this note was to determine whether both familiarity and social desirability as stimulus variables are related to subjects' response selections in a choice task. A second purpose was to determine whether the responses of neurotic and control groups were differentially related to these variables.

The findings showed that for control subjects, both relative familiarity and social desirability were significantly related to preferences for alternatives when the other variable was held constant. With neurotic subjects, social desirability did not contribute significantly to the variance in their preferences for alternatives when familiarity was partialled out. Familiarity and response preferences were significantly correlated when social desirability effects were controlled.

Differences between the obtained partial *r*'s were not significant; nevertheless, they were in the relative direction expected from Mowrer's proposition that neurotic people are less able to apply social values than nonneurotic people.

REFERENCES

1. BRUNER, J. S., & POSTMAN, L. Emotional selectivity in perception and reaction. *J. Pers.*, 1947, **16,** 69–77.
2. EBEL, R. L. Estimation of the reliability of ratings. *Psychometrika*, 1951, **16,** 407–424.
3. JENKINS, N. Affective processes in perception. *Psychol. Bull.*, 1957, **54,** 100–128.
4. MCGINNIES, E. M. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56,** 244–251.
5. MOWRER, O. H. *Learning theory and personality dynamics.* New York: Ronald Press, 1950.
6. POSTMAN, L. J. The experimental analysis of motivational factors in perception. In *Current theory and research in motivation, a symposium.* Lincoln, Nebr.: Univer. Nebraska Press, 1953. Pp. 59–107.
7. POSTMAN, L. J., BRONSON, W. S., & GROPPER, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1953, **48,** 215–224.

8. POSTMAN, L. J., & CRUTCHFIELD, R. S. The interaction of need, set, and stimulus-structure in a cognitive task. *Amer. J. Psychol.*, 1952, **65**, 196–217.

9. SOLOMAN, R. L., & HOWES, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256–270.

10. WATSON, R. E., PRITZKER, L., & MADISON, P. Hostility in neurotics and normals. *J. abnorm. soc. Psychol.*, 1955, **50**, 36–40.

11. WHITTAKER, EDNA M., GILCHRIST, J. C., & FISCHER, JEAN W. Perceptual defense or response suppression? *J. abnorm. soc. Psychol.*, 1952, **47**, 615–623.

Received July 18, 1957.

THE EFFECT OF RECENT AND REMOTE FRAMES OF REFERENCE ON TEMPORAL JUDGMENTS OF SCHIZOPHRENIC PATIENTS[1]

ALVIN D. WEINSTEIN, SANFORD GOLDSTONE, AND WILLIAM K. BOARDMAN

Baylor University College of Medicine and Houston State Psychiatric Institute

IT HAS long been recognized that the so-called *absolute* judgment in a discrimination situation is influenced by the range of magnitudes presented in the experimental context. Subjective scales for such qualities as loudness-softness, brightness-dullness, and longness-shortness are relative in nature, depending upon subject's (*S*'s) frame of reference when the discrimination is requested. Subjective scale shifting in the direction of extreme stimuli in an experimental series has been called the *anchor effect* (5). It has been suggested (2) that *S* selects the end stimuli in a series as standards for making judgments and that this contextual frame of reference accounts for the anchor effects.

Schizophrenic *S*s were found (7) to differ from normals in their absolute judgments of short auditory durations when a method was used that was sensitive to anchor effects (4) as well as reflecting changes in temporal judgment as a function of development (9) and drug states (1, 3). With a measure of *S*'s concept of a clock second or Second Estimation Point (SEP) obtained by presenting alternating ascending and descending series of tones to *S*, who judged each sound as *more* or *less* than 1 second, schizophrenics overestimated the clock second to a greater extent than did normals when the initial duration or anchor was 1 second (7). The judgments of normals (4) were influenced by the magnitude of the anchor stimulus. If the first or anchor duration was short (.1 sec.), *S* estimated a shorter duration as equal to a clock second than when the initial stimulus was longer (1.0 or 2.0 sec.). When a long anchor series followed a short one or a short one followed a long anchor series, there was an interaction between initial anchor-determined frame of reference and final anchor-determined frame of reference. Apparently, normal *S*s incorporate past experience with other anchors in making current judgments.

The present study investigated the effects of a long anchor (more than 1 sec.) and a short anchor (less than 1 sec.) upon estimations of a clock second (SEP) by schizophrenics. Is the SEP or subjective second influenced by anchors in a manner comparable with normal *S*s, or does the patient rigidly stick to an autistic concept and remain isolated from contextual influences? In addition, the independence of anchor effects was studied by determining the interaction between long and short anchor series in the same experimental context. Is a residual effect of a remote anchor observed as in the normal group, or is one anchor-determined frame of reference relinquished upon the appearance of a new and different anchor? Finally, this study examined the previous finding (7) that schizophrenics overestimate duration to a greater extent than normal *S*s. Do patients have a shorter SEP under all anchor conditions?

METHOD

Subjects. Forty-six schizophrenic patients[2] were divided into two equal groups. Group 1 contained 14 men and 9 women, with a mean age of 36 years. Group 2 contained 14 men and 9 women, with a mean age of 34 years. All patients hospitalized were in most instances considered chronic. Diagnoses by the hospital staff were accepted for the experiment. Patients consecutively available who met the requirement of at least high school education or its equivalent were used.

Eighty healthy college students were similarly divided into two groups. The first contained 20 men and 20 women, with a mean age of 25 years. The second contained 20 men and 20 women, with a mean age of 23 years.

Apparatus. A decade interval timer activated an audio-oscillator that delivered through headphones a constant amplitude tone of 725 cps for durations of

---

[1] Supported in part by a grant from the National Institute of Mental Health, USPHS M-1121.

[2] We are indebted to S. A. Hoerster, of the Austin State Hospital and the Psychology Department of the University of Houston, for making *S*s available.

TABLE 1

THE MEAN LOG SEP, *SD*, AND GEOMETRIC MEAN SEP (ANTILOG MEAN LOG SEP) UNDER ALL ANCHOR CONDITIONS FOR PATIENT AND CONTROL GROUPS

	Group 1 Short (0.1 sec.) Followed by Long (2.0 sec.) Anchor Series		Group 2 Long (2.0 sec.) Followed by Short (0.1 sec.) Anchor Series	
	Control	Patient	Control	Patient
Block 1, Initial Anchor Series				
Anchor Duration	0.1 sec.	0.1 sec.	2.0 sec.	2.0 sec.
N	40	23	40	23
Mean Log SEP	$\bar{1}.6452$	$\bar{1}.3769$	$\bar{1}.8302$	$\bar{1}.6537$
SD Log SEP	.175	.346	.270	.485
Geo. Mean SEP	.442	.238	.671	.450
Block 2, Final Anchor Series				
Anchor Duration	2.0 sec.	2.0 sec.	0.1 sec.	0.1 sec.
Mean Log SEP	$\bar{1}.7663$	$\bar{1}.7123$	$\bar{1}.7613$	$\bar{1}.3241$
SD Log SEP	.188	.354	.252	.522
Geo. Mean SEP	.584	.516	.577	.211

.1 to 9.9 sec. in .1-sec. increments, with an accuracy of 5% and a repeat accuracy of 1%. An additional circuit in the timer provided durations from .02 to .09 sec. in .01 sec. increments. The reliability in this range was not as great as for the longer durations, but the arrangement allowed a gross extension of the lower end of the stimulus scale.

Procedure. *S*s were instructed to report whether each tone heard was more or less than 1.0 sec. A method of limits technique with a .1-sec. step interval was used. Twenty alternating runs of ascending and descending durations were initially presented. If *S* reported *more* to the first duration presented, the first run was descending, terminating when *S* reported *less* to three consecutive stimuli. The following run was ascending, beginning one step interval below the shortest duration of the previous descending series and terminating when *S* reported *more* to three consecutive stimuli. Another descending run was then initiated, beginning one step interval above the longest duration of the previous run, and so on. If *S* responded *less* to the initial stimulus, the first run was ascending.

Between .1 and 9.9 sec., step intervals of .1 sec. were used, and a step interval of .01 sec. was employed between .02 and .1 sec. The first 10 of the initial 20 runs were considered to represent a period of stabilization and were not included in any statistical treatment. Subsequent series consisted of only 10 runs.

Patients and normals in Group 1 were tested in two blocks as follows: Block 1 consisted of 20 alternating runs of ascending and descending durations, starting at .1 sec. (short anchor). Following a 1-min. rest, Block 2 was given, in which there were 10 alternating runs of descending and ascending durations starting at 2.0 sec. (long anchor). Patients and normals in Group 2 were similarly tested in two blocks, but the first block started at 2.0 sec. and the second at .1 sec.

Percentages of reports of *less* for each duration of the four blocks in the two groups were plotted on normal probability paper and a straight line visually fitted. The Second Estimation Point (SEP) or duration which *S* reported *more* or *less* 50% of the time and the slope or standard deviation of *S*'s judgments were derived from these plots.

RESULTS

SEP data were log normally distributed, and the appropriate transformations were made. Table 1 shows the mean log SEPs, their *SD*s, and the geometric means (antilog mean log SEP). In the interest of clarity, results will be described in terms of the geometric means. Differences between the mean log SEPs of the various anchor conditions for the patient and control groups were examined by t test, taking heterogeneous variance into account when appropriate.

For patients, the .1 sec. initial (Group 1) anchor series provided a significantly ($P < .005$; $t = 5.53$) smaller geometric mean SEP (.238 sec.) than the 2.0 sec. initial (Group 2) anchor series (.450 sec.). For the controls, the .1 sec. initial anchor series also provided a significantly ($P < .005$; $t = 3.70$) smaller geometric mean SEP (.442 sec.) than the 2.0-sec. initial anchor series (.671 sec.). The schizophrenic geometric mean SEP was significantly smaller ($P < .005$; $t = 3.47$) than that obtained from the controls for the initial .1-sec. anchor condition.

These results suggest that schizophrenic *S*s tend relatively consistently to overestimate the magnitude of a clock second to a greater extent than controls, substantially confirming the earlier report (7). The patient group demonstrated the anticipated initial anchor effect; the short anchor "pulled" the SEP down, while the long anchor "pulled" up the SEP.

In order to study anchor interaction, the short initial anchor group was given a final long anchor series, and the long initial anchor group was given a final short anchor series. For the control group, the difference noted between the mean log SEPs produced by the long and short initial anchors disappeared under this final "anchor reversed" condition. The geometric mean SEP obtained with a 2.0-sec. final anchor series, preceded by a .1-sec. initial anchor (.584 sec.), did not differ ($t = 0.10$) from that obtained with a final .1-sec. anchor preceded by a 2.0-sec. initial anchor series (.577 sec.). The SEPs obtained under the final anchor conditions reflect a compromise between the initial anchor and the anchor of the immediate series. Healthy *S*s modified their frames of reference in accordance with the context of the situation with a measurable residual effect of recent past frames of reference.

This interaction was not observed in the patients. The downward pulling effect of the .1-sec. final anchor was not influenced by any residual effect of the initial 2.0-sec. anchor, and the upward pulling effect of the 2.0-sec. final anchor was not influenced by the initial .1-sec. anchor. The geometric mean SEP for the final 2.0-sec. anchor series (.516 sec.) was significantly ($P < .005$; $t = 3.77$) longer than the .1-sec. final anchor SEP (.211 sec.). Instead of demonstrating a compromise between initial and final anchor as noted in the healthy groups, the schizophrenics produced a final SEP that corresponded to the demand value of the anchor independent of past experience with another anchor.

Median coefficients of variation (slope/SEP) of 17% and 16% were obtained for the schizophrenic 2.0-sec. and .1-sec. anchor groups respectively, as compared with 16% in both cases for the control groups. There was thus no evidence of differences in intrasubject variability of judgment among groups.

DISCUSSION

Confirming an earlier report (7), the findings strongly suggest that the schizophrenic patient is likely to overestimate the duration of a clock second to a greater extent than normal controls under all anchor conditions. No satisfactory explanation of this overestimation is readily available, although a general relative slowing of world time seems consistent with the passivity and autism accompanying schizophrenia. Presumably, the mental activity involved in fantasy proceeds at a more rapid rate than comparable, externally directed motor activity. Less clock time is necessary to *think* an event than to *do* an event. When the reference standard is the clock and fantasy activity dominates behavior, external events necessarily appear slower by comparison. The experience of time, then, must be considered in terms of internal frames of reference.

Since schizophrenics seemed influenced by immediate situational or contextual factors in a manner comparable with controls, the question remains of the residual effect of past experience in similar contexts on temporal judgments. This question was dealt with in the third part of this study, which attempted to investigate the interaction of recent with remote contextual influences by changing anchor during the same experiment. When long and short anchors were reversed, the normals demonstrated an anchor interaction, giving judgments which incorporated the past anchor-determined frame of reference; the schizophrenics, however, ignored past experience with anchors, judging solely in terms of the demand value of the new anchor. It is as though the patients were unable to use past anchor-determined frames of reference. These results are similar to the findings of reaction time studies (6, 8) showing that schizophrenic patients do not maintain a mental set as well as normals. Control *S*s profit by a constant interval preparatory signal, while patients fail to profit by this knowledge. In analogous terms, the schizophrenic is deficient in taking advantage of past experience in conceptualizing current frames of reference. His frame of reference is ahistoric; he is at the mercy of imminent stimulation. This tendency to respond to the demand value of immediate environmental anchors, without taking into account internalized anchor-determined frames of reference which might modify judgment, appears to reflect the passivity of the schizophrenic patient.

The effects of drug therapy and remission upon the SEP and anchor interaction of schizophrenic *S*s remain as fruitful areas of investigation.

SUMMARY

Forty-six schizophrenic patients were compared with 80 control *S*s with respect to their estimation of one clock second's duration under long and short anchor conditions, where anchor represented the extreme duration in a series. In addition, the interaction of recent and more remote anchors was studied. The following findings are reported:

1. Patients had a smaller Second Estimation Point (SEP) than did controls under all anchor conditions.

2. The influence of immediate anchors on temporal judgments seemed similar for patients and controls, in accord with the usual pulling effect of anchors.

3. The judgments of the control *S*s were influenced by remote anchors in the same experimental context, whereas schizophrenic *S*s responded only to the pulling effect of immediate anchors.

The relation of these findings to the psychopathology of schizophrenia and frame-of-reference theory is briefly discussed.

REFERENCES

1. BOARDMAN, W. K., GOLDSTONE, S., & LHAMON, W. T. The effects of LSD on the time sense of normals: A preliminary report. *Arch. Neurol. Psychiat.*, 1957, **78**, 321-324.
2. ERIKSEN, C. W., & HAKE, H. W. Anchor effects in absolute judgments. *J. exp. Psychol.*, 1957, **53**, 312–314.
3. GOLDSTONE, S., BOARDMAN, W. K., & LHAMON, W. T. Effect of secobarbital, dextroamphetamine and placebo upon apparent time. *Brit. J. Psychol.*, in press.

4. GOLDSTONE, S., LHAMON, W. T., & BOARDMAN, W. K. The time sense: Anchor effects and apparent duration. *J. Psychol.*, 1957, **44**, 145–153.

5. GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1954.

6. HUSTON, P. E., & SINGER, M. M. Effect of sodium amytal and amphetamine sulfate on mental set in schizophrenia. *Arch. Neurol. Psychiat.*, 1945, **53**, 365–369.

7. LHAMON, W. T., & GOLDSTONE, S. The time sense: Estimation of one second durations by schizophrenic patients. *Arch. Neurol. Psychiat.*, 1956, **76**, 625–629.

8. RODNICK, E. H., & SHAKOW, D. Set in the schizophrenic as measured by a composite reaction time index. *Amer. J. Psychiat.*, 1940, **97**, 214–225.

9. SMYTHE, E., & GOLDSTONE, S. The time sense: A normative genetic study of the development of time perception. *Percept. Motor Skills*, 1957, **7**, 49–59.

Received July 31, 1957.

PERSONALITY DIFFERENCES IN THE CAPACITY FOR VERBAL RECALL

JOSEPH ADELSON
University of Michigan
AND JOAN REDMOND
Boston, Massachusetts

WHEN we think of the topic "psychoanalysis and memory" we are likely to bring to mind Freud's theory of repression; and we are also apt to think of that curious scatter of experiments—some of them knowing and clever, some of them innocent of understanding—which have striven to confirm or confute Freud's observations. Many of these studies concentrate on the stimulus: they ask whether a certain type of stimulus—"disturbing," "unpleasant"—can, by evoking anxiety, bring into play the ego's defenses so as to inhibit or distort recall. Other studies center on the experimental subject: by inducing a momentary state of distress, they seek to bring about some disturbance in recall.

Our research approaches psychoanalysis and memory from a somewhat different perspective. It focuses on the problem of individual differences in the *capacity* for recall. The topic is much neglected in an otherwise extensive tradition of research; there have been few studies of differences in recall capacity; and there appear to have been none that look to personality as a source of such differences. At a first glance, psychoanalytic theory seems to offer no explicit statement on the problem, but a closer examination yields this hypothesis: individuals fixated at the late anal phase (the so-called anal retentives) have a greater ability to recall verbal material than those fixated at the early anal phase (anal expulsives).

Surely a far-fetched idea; but it may gain plausibility if we take the retentive-expulsive dimension to specify, not psychosexual fixation points per se, but differing forms of ego organization.[1] To begin with the anal retentive: (*a*) He has developed fairly stable techniques for coping with aggressive impulses. The dominant defenses are reaction-formation, undoing, and the various forms of isolation (including intellectualization). The isolation defenses are especially relevant here: their use gives a systematic and orderly cast to the thought process; verbal stimuli can be effectively organized and codified; when counter-cathexes are directed against affects, ideation proper remains undisturbed. (*b*) His defensive style is characterized by a peculiar absorption in words and concepts. As Fenichel puts it: "He flees from the macrocosm of things to the microcosm of words" (1, p. 295). Cathexes are directed towards language, ideas, the thinking process itself.

These processes occur among (indeed, partly define) the obsessional neuroses. In the severely obsessional personality, however, we generally find that cognitive functioning is seriously burdened by the strenuous demands of defense; among "normal" retentives, cognitive style is in the obsessional direction—systematic and orderly—yet is spared the brittleness and rigidity, the vulnerability to breakthrough that we see among the clinically obsessional.

Far less is known about "anal expulsive" normals, and so we must offer here a more hesitant formulation. We take it that in this group, too, character structure has developed out of the struggle between parent and child on the issue of aggression. The anal retentive tames his hostile drives by turning against them, or by isolation; the expulsive person shows a less thorough, less efficient regulation of these impulses. Aggression continues to be discharged, but in a muted, modified way, the dominant defenses being aim-inhibition and displacement. The original aims—aggression and disorder—are, so to speak, blunted, and find a modulated continuation in messiness, rebelliousness, impulsiveness.

---

[1] The discussion will follow a line of approach suggested by Fenichel (1, pp. 295–300) and Rapaport (3, pp. 622–625).

The intrapsychic system, then, shows a *relative* failure to neutralize aggressive drives or to bind them adequately through countercathexes. A persistent imbalance between impulses and defense produces sporadic inefficiencies in certain types of cognitive performances. Leakages of aggression elicit anxiety and so bring about momentary disruptions of attention and concentration.

We are now in a position to turn to the problem of recall. Among retentives, the use of isolation permits a heightening of attention to external stimuli; unwelcome impulses and effects, which would disturb attention, are warded off. Equally important, we find a tendency to hypercathect words, an increased alertness to them. In concert, these processes produce a peculiar efficiency in the apprehension of verbal stimuli. Expulsives, on the other hand, are characterized by defensive processes that lead to a decreased capacity for verbal recall; the manner of handling impulses and affects brings about momentary disturbances of attention and concentration.

METHOD

Subjects. The sample consisted of 61 college women, all of them first-year students at Bennington College.

The Criterion for Anality. The Blacky Test was used to determine the type of anal fixation. Several weeks prior to the experiment itself, the test had been administered on a group basis to all entering students at the college. An *S* was designated "Expulsive" if she received a score of ++ (very strong) or + (fairly strong) on Anal Expulsion, and a score of 0 on Anal Retention; the reverse was true for the scoring of retentiveness. *S*s receiving scores on both dimensions were not included. An *S* was rated "Neutral" if she received 0 scores on both dimensions. The experimental group was composed of 32 expulsives, 18 retentives, and 11 neutrals.

The Stimuli. Two prose passages were used, one with innocuous, the other with disturbing content. We wanted to determine whether the hypothesized differences in recall would be present only when the material was of a nature to evoke defenses, or whether they would also be found with relatively neutral material.

a. "Disturbing" passage. Here we wanted a passage which would be threatening enough to elicit defenses, yet sufficiently decorous to permit its presentation to young ladies in a classroom setting. So we devised a passage which stated, rather crudely, Freud's theory of psychosexuality. It follows:

> Psychoanalysis is the creation of Sigmund Freud, who was born in 1856 and died in 1939. Beginning with the attempt to cure neurotic patients, Freud developed one of the most significant systems for understanding human behavior. His psychology has had widespread influence to the extent that it has completely revised the previous understanding of human nature. At the center of his system is the idea that early childhood experiences influence later personality. This is called the Theory of Psychosexual Development, the major features of which follow:
>
> During the first oral stage, the child receives intense pleasure sucking the mother's breast. There is a desire to eat everything and resistance to being deprived of this pleasure when weaned.
>
> During the second oral stage, intense pleasure is received through using the teeth to bite and devour, desiring to chew and gnaw at the mother, with fear of punishment in return.
>
> During the anal stage, pleasure is obtained through retaining bowel movements and through eliminating them. The child resists being trained to be neat and clean, and would rather smear his feces.
>
> During the oedipal stage, the female child feels sexually attracted to her father, is jealous of her mother for possessing him, and wishes for the death or disappearance of her mother.
>
> Between the third and fifth year, children experience intense pleasure through masturbation, by playing with their sexual organs. Along with this there is a strong fear of being caught and punished.
>
> During this same period, female children become aware that they lack a penis. They feel inferior, believing that its absence is a punishment, and are jealous of the male's sexual organ.
>
> Children sometimes feel intense hostility and jealousy toward a sibling. They feel that they are being neglected in favor of the other child, and wish for the sibling's death or disappearance.

b. "Innocuous" passage: Here we wanted didactic, colorless prose, and so chose the following passage:

> Undoubtedly the most famous brick house in seventeenth century New England was that built for Peter Sergeant in Boston. Sergeant, who had come to Boston in 1667, became a wealthy merchant and later served as a judge and a member of the Governors Council. The land he bought on October 21, 1676, a large tract extending halfway from Washington to Tremont Street, included a fine garden and orchard, stable, and coach house. The big house which he built with brick walls about 2 feet thick, was completed by 1697. The Sergeant House was early accustomed to society. Governor and Lady Bellomont visited it for several months in 1697, and Sergeant's third wife, whom he married in 1707, was Lady Phips, widow of the former Governor Sir William Phips. After Sergeant's death in 1714 the mansion was purchased by the province (in 1716) to serve as a residence for the royal governors. It was thenceforth known as the Province House.

c. Scoring. Because of differences in the sentence structures of the two passages, we used a separate scoring system for each. The "disturbing" passage

TABLE 1

DIFFERENCES BETWEEN EXPULSIVES AND RETENTIVES IN VERBAL RECALL

(Expulsives $N = 32$; Retentives $N = 18$)

Tests	Group	*M*	*SD*	*t*	*P*
Innocuous passage–Immediate recall	Expulsive Retentive	19.9 28.3	12.84 12.08	2.25	.05
Innocuous passage–Delayed recall	Expulsive Retentive	9.9 15.1	6.00 8.66	2.60	.02
Disturbing passage–Immediate recall[a]	Expulsive Retentive	21.6 28.7	9.59 12.29	2.68	.02
Disturbing passage–Delayed recall[a]	Expulsive Retentive	11.1 19.3	5.92 8.94	3.66	.001

[a] The mean scores for the disturbing passages refer to thought units.

had been written so as to permit its being scored by phrases, or thought units; *S* was given one point for each phrase precisely or substantially correct. In the scoring of the "innocuous" passage, one point was given for each word (except for conjunctions and prepositions) correctly reproduced; dates and numbers accurately recalled were given two points. Since the scoring of the "innocuous" passage was more or less mechanical, requiring only the counting of words, no study of interjudge agreement seemed necessary. For the scoring of the "disturbing" passage, scoring criteria were first established for each phrase; then a sample of the protocols was scored separately by two judges. Agreement between them was nearly complete and formal testing of interjudge agreement was therefore omitted.

Administration. The recall tasks were administered to five classes in freshmen literature which met at the same hour. The experimenter introduced the task in these words: "Read these papers carefully twice. You will be asked questions on them later." *S*s were then given the passages to read. The passages were printed on separate sheets of paper. The classes were allowed about ten minutes to read the material. The passages were then collected, writing paper distributed, and *S*s told: "Now write down all you can remember about what you have just read." This was a test of immediate recall. We were also interested in learning what differences, if any, would be present in delayed recall. One week later, the experimenters therefore returned to the classes and asked the *S*s to reproduce all they could now remember about the passages.

There were 78 *S*s in the five classes; all are included in the analysis, except those with Blacky scores on both anality dimensions, and those who were absent from either or both of the class meetings when the recall testings were done.

RESULTS AND DISCUSSION

Anal Retention vs. Expulsion and Recall

In Table 1, we find the results for both passages and for immediate and delayed recall. Retentives show a significant superiority in reproduction of both types of material; their advantage persists over time.

We may gain a more graphic impression of the differences if we look at the data from the point of view of quartile extremes. For example, when we consider the distribution of scores for "disturbing passage—immediate recall," we find that retentives, who make up 36 per cent of the total sample, contribute 69 per cent of the scores in the top quartile and only 9 per cent in the bottom quartile. Roughly equivalent distributions are found for the other passage and other testings.

We made no specific hypothesis about the 11 neutral *S*s; however, we wanted to determine the recall performance of this group. If we take these *S*s to represent average functioning in recall, we would have some basis for judging whether the expulsive-retentive differences are due to the superior recall capacity of retentives, the inferior capacity of expulsives, or both. We found that for all four comparisons the neutral group is located midway between the other two. This finding suggests that both the retentive and expulsive *S*s occupy extreme positions in recall performance.

Other Possible Sources of the Differences

a. Other psychosexual dimensions. We next considered whether differences in recall are associated exclusively with anality; could it be that anality is one of several psychosexual dimensions where we would find similar differences? To test this possibility, we looked for recall differences (both passages, both testings) on all of the other Blacky dimensions; that is, we compared the recall scores of high and low groups on Oral Sadism, Oral Eroticism, Oedipal Intensity, and so on; we also compared high and low groups on Total Pregenitality. Of 40 differences tested, only two were significant at the .05 level: the high group on Oedipal Intensity has better recall scores for the "innocuous passage—delayed recall"; the low group on Ego-Ideal has higher recall scores on "disturbing passage—immediate recall." These results (two of 40 at the .05 level) are what we would expect by chance alone.[2]

It is interesting to note that no Blacky group receives recall scores (on each of the four passage-testing comparisons) as high as those received by the retentives; and only one Blacky group has a mean recall score as low as those received by the expulsives.

---

[2] The tables which report these and other results described here may be found in Redmond (4).

b. Intellectual capacity. Another possible explanation of the results is that expulsives and retentives are intellectually unequal, and that the differences in recall performance are due to intellectual differences associated with the type of anal fixation. To check this, we compared the SAT scores (V and M) of both groups; these were available for most of our *S*s. For both V and M there are no significant differences between the groups; indeed, their means are almost identical.

A Partial Replication

An unpublished study by Nahin (2) supports some of these findings. Nahin's research was on a somewhat different topic, but a portion of her study duplicates some elements of the present one. The criterion for anality was the Blacky Test; *S*s were Bennington College freshmen. However, she used only a "disturbing" prose passage (basically equivalent to the one reported here) and tested only for immediate recall. She found significant differences between expulsive and retentive *S*s.

Nonverbal Recall

What about the recall of nonverbal stimuli? Should we expect to find expulsive-retentive differences here as well? The question is relevant to some parts of the theory underlying this research. The retentive *S* is presumed to be sensitive to words, at least in some respects: he tends to reduce the anxiety inherent in "things" by capturing them verbally, by transposing them into the safer, more tractable realm of language: taming by naming, let us say. There develops, then, a peculiar efficiency in certain (though surely not all) verbal performances, as this study has suggested. Now one might wonder whether this capacity is operative only when the stimuli are verbal (or can be made so by being "named" or "labelled"), or whether we are dealing with a more general disposition.

We cannot, unfortunately, provide a satisfactory answer to the question. Although we carried out a study of the recall of nonverbal stimuli, circumstances did not allow a precise enough duplication of the conditions of the verbal recall research. It is reported here briefly.[3]

A set of 20 abstract line drawings were developed. The pictures were simple enough in design to permit ready reproduction by the experimental *S*s, and were so drawn that they did not resemble common forms or objects; consequently, they could not be "labelled." In the experiment, the drawings were projected on a screen, one at a time, for 15 seconds. *S*s were students in a large lecture class at the University of Michigan, all of whom had earlier been tested for expulsion-retention. Except for slight and necessary changes in wording, the instructions were the same as in the verbal recall study. After the drawings had been exposed, the group was asked to reproduce them. The reproductions were scored by a six-category scheme (e.g., perfect reproduction; one slight error, etc.). Male and female *S*s were treated separately, since they distribute differently on expulsion-retention; there were 33 female retentives and 22 expulsives, 21 male expulsives and 10 retentives. Expulsion-retention differences, within sex, were tested for each of the six scoring categories and for various combinations of categories. No significant differences were found; in fact, recall scores throughout are highly similar.

But the variations in experimental procedure are too marked to allow us to treat the two recall studies as comparable. In all likelihood, recall performance can be significantly influenced by differences in motivational conditions, experimental set, testing atmosphere, and so on. It seems the most judicious course to take the findings on nonverbal recall as, at best, suggestive.

Discussion

The study has demonstrated that under certain conditions we can obtain personality differences in verbal recall. Yet all of our experience in the area of personality and cognition warns us to caution in generalizing the findings. The results may very well be tied to a particular set of experimental conditions: change the stimuli, in length or content; change the experimental instructions; change the method of testing recall, from reproduction to recognition, as an example; use a different sampling of *S*s; make any of these changes and, for all we know, the differences may fail to appear. Our understanding of the expulsion-retention variable and of the influence of personality on recall is still too uncertain to allow us to extrapolate the findings with any sense of confidence.

There are, then, a great many occasions for speculation offered by the research. The stimuli, the details of administration, the experimental atmosphere (especially the role of anxiety in producing the results)—each of these gives us an opportunity to reconsider the meaning of the experiment. But perhaps the most critical question to be raised concerns the expulsion-retention variable itself.

In formulating the central hypothesis of the research, we took the position that recall differences are based on capacity differences; equivalent motivation was assumed for both groups of *S*s. The assumption, however, is open to challenge: can we explain the findings equally well as arising from motivational differences between expulsives and retentives?

---

[3] The authors are grateful to John Hirtzel for carryout the analysis of the data.

A considerable body of unpublished data makes this a highly credible alternative. Anal retentives, it appears, show a marked disposition towards compliance and conformity, especially so in the presence of authority; anal expulsives are distinguished by an edgy, often sullen independence. Thus we can well imagine the retentives, solemn, *bürgerlich*, eager to do well, bending to the experimental task with dedication and gravity; and it is quite as easy to see the expulsives, skeptical or diffident or rebellious, giving the task only cursory attention.

There is, at this moment, no way of choosing between the motivational and cognitive interpretations of the findings; they are equally plausible. If the nonverbal recall experiment had been comparable to the earlier study, then we would have some reason for believing that motivational differences are not the source of the recall differences. (Even so, it could be argued that nonverbal stimuli do not engage motivation or interest to the same degree that words do.) As it is, the question must remain open, awaiting further research.

The position we take on this issue—motivation vs. capacity—will determine whether we view the recall differences as a function of differences in acquisition, or retention, or performance. If a motivational factor is involved we can regard it as influential at any or all points between stimulus presentation and reproduction. The contrast between eager and indifferent attitudes towards the task might operate in the amount of attention given the stimuli, or in the intention to retain what was learned, or in the degree of interest in reproducing the stimuli.

Our own approach is derived from the cognitive concepts of psychoanalytic theory, and so we interpret the findings as an outcome of variations in cognitive style. From this viewpoint, one is led to stress differences in the capacity for attention and concentration, which would operate with particular force during the process of acquisition. We would then see the experiment as a study of incidental learning, rather than retention or performance.

SUMMARY

It was hypothesized that "anal retentive" individuals have a greater ability to recall verbal material than "anal expulsive" subjects. The hypothesis was derived from an analysis of differences in ego organization between the two groups.

The subjects were 61 first-year college women. The criterion of "anality" was the Blacky Test; there were 32 expulsives, 18 retentives, and 11 neutrals. Subjects were asked to read two prose passages, each of several hundred words. One passage discussed sexual and aggressive themes and was considered "disturbing" in content; the other was an "innocuous" description of colonial architecture. Subjects were asked to reproduce the passages immediately after presentation and again one week later.

Retentives showed a significant superiority over expulsives in the reproduction of both passages, immediately and one week later. Neutrals scored midway between the two extreme groups on all four testings. There were no differences in intellectual capacity between retentives and expulsives, as measured by the SAT., V, and M. None of the other Blacky dimensions is associated with recall differences. The findings have been partially corroborated in another study. A study of the recall of nonverbal stimuli showed no differences between expulsives and retentives; since the research methods of this study differed in some possibly critical respects from the investigation of verbal recall, the findings are offered as suggestive rather than definitive.

REFERENCES

1. Fenichel, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.
2. Nahin, Barbara. Psychosexuality and memory. Unpublished thesis, Bennington College, 1953.
3. Rapaport, D. *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951.
4. Redmond, Joan. Anality and memory. Unpublished thesis, Bennington College, 1954.

Received June 17, 1957.

VERBAL CONDITIONING AND ANXIETY

ARNOLD H. BUSS
University of Pittsburgh

AND IRMA R. GERJUOY
Carter Memorial Hospital

TAFFEL (3) reported a relationship between verbal conditioning and scores on the Taylor Manifest Anxiety Scale (MAS). The task consisted of selecting one of six pronouns and pairing it with a verb to make up a sentence. Whenever first person pronouns were used, *E* said "Good." It was found that High Anxious and Medium Anxious *S*s conditioned (their frequency of first person responses increased), whereas Low Anxious *S*s did not.

The present data were gathered in the course of a study of verbal conditioning (1) in relation to different kinds of reinforcers. Both the task and the MAS cut-off scores were identical to Taffel's. The only difference between the studies was in *S*s. Taffel employed male Veterans Administration psychiatric inpatients; *S*s here were 45 state hospital psychiatric patients of both sexes. It would seem that the studies are basically comparable.

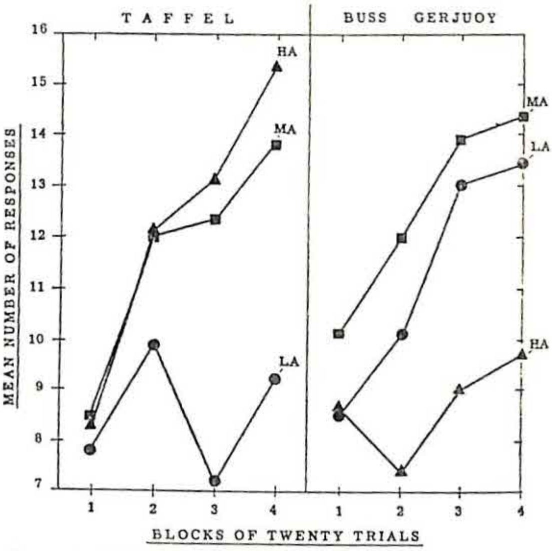

FIG. 1. ACQUISITION CURVES OF HIGH ANXIOUS, MEDIUM ANXIOUS, AND LOW ANXIOUS GROUPS IN TAFFEL'S AND THE PRESENT STUDY

METHOD

Task. There were 80 3 x 5 cards, each with a verb in the past tense and six pronouns. *S* was instructed to make up a sentence, using the verb and one of the six pronouns. The order of pronouns differed from card to card, and each card had a different verb.

Procedure. During the first 20 trials, *E* merely recorded *S*'s choice of pronouns. At the end of 20 trials, one group of *S*s was told that they would now receive chips on some of the trials and that the chips could be turned in afterward for either cigarettes or candy. During Trials 21–80, *E* gave a chip after every sentence that began with "I" or "We."

The remaining *S*s were given no instructions at the end of the first 20 trials. During Trials 21–80, *E* said "Good" after every sentence that started with a first person pronoun.

The Taylor MAS was administered to each *S*, and Taffel's cut-off scores were used to identify three anxiety groups: Low Anxious (LA) 0–24; Medium Anxious (MA) 25–34; and High Anxious (35–50).

RESULTS AND DISCUSSION

Because differences between the Cigarettes-Candy group and the Good group were small and random, the data were combined. The distribution of *S*s in the anxiety groups was as follows: males, LA—13, MA—7, and HA—1; females, LA—9, MA—7, and HA—8. Before the effects of anxiety were evaluated, sex differences within the anxiety groups were tested. A summary score for each *S* was obtained by subtracting the frequency of first person responses for the first 20 trials from the frequency for the last 20 trials. The differences between men and women could be tested only in the LA and MA groups because there was only one male *S* in the HA group. For the LA group, t was 1.65, which yields a p of more than .10 for 20 df. For the MA group, t was less than 1.0. In the absence of statistically significant differences in the LA and MA groups, it appears unlikely that such differences would have occurred in the HA group had there been enough *S*s to make the test. Therefore the males' and females' data were combined.

The acquisition curves for the three anxiety groups,together with Taffel's comparable curves, are presented in Fig. 1. Inspection of the left half of Fig. 1 reveals that Taffel's HA and MA groups conditioned, but his LA group did not. In the present study, the relationships among the anxiety groups are somewhat different, as shown in the right half of Fig. 1. Here the LA curve rises, whereas the HA curve remains relatively flat. Thus, Taffel's results were not replicated.

These negative findings are consistent with those of Daily (2), who studied verbal conditioning, using college *S*s and different MAS cut-off scores. He found no significant difference between his high anxious and low anxious *S*s.

When the results of the present study are taken together with Taffel's and Daily's, two conclusions seem tenable:

1. There is a relationship between Taylor MAS scores and verbal conditioning, but it holds only for a selected population of Veterans Administration male psychiatric inpatients.

2. There is no relationship between MAS scores and verbal conditioning, and Taffel's positive finding may be attributed to random fluctuations in sampling.

REFERENCES

1. Buss, A. H., Gerjuoy, Irma, & Zusman, J. Verbal conditioning and extinction with verbal and nonverbal reinforcers. *J. exp. Psychol.*, 1958, **56**, 139–145.
2. Daily, J. M. Verbal conditioning without awareness. Unpublished doctoral dissertation, State Univer. of Iowa, 1953.
3. Taffel, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.

Received August 30, 1957.

HYSTERICS AND DYSTHYMICS AS CRITERION GROUPS IN THE STUDY OF INTROVERSION-EXTRAVERSION: A REPLY

H. J. EYSENCK

Institute of Psychiatry, University of London

In their interesting study of the validity of the Maudsley Personality Inventory, Sigal, Star, and Franks (3) come to an incorrect conclusion because of several logical errors which deserve brief discussion.

The hypothesis underlying the use of hysterics, psychopaths, and dysthymics as criterion groups for extraversion-introversion derives, of course, from Jung, not, as Sigal et al. state, from myself. Jung maintains "that much the most frequent neurotic disorder of the extraverted type is hysteria"; on the other hand, speaking of the introvert, he maintains that "his typical neurotic disorder is psychasthenia," or, as I would prefer to call it, dysthymia. Two consequences would seem to follow from this hypothesis. In the first place, among neurotics, those of a dysthymic pathology should be more introverted than those of a hysteric or psychopathic pathology. (What is now called "psychopathy" is traditionally considered as part of the complex of disorders diagnosed as "hysteria"; I have documented this statement in [2, ch. 6].) This is the crucial deduction on which hinges the usefulness of these nosological groups for the purpose of questionnaire validation.[1] The second deduction to be made from Jung's hypothesis is that normal groups should be intermediate between the hysteric and psychopathic groups on the one hand, and the dysthymic groups on the other. The truth or falsity of the second proposition, while of considerable theoretical interest, is not crucial for the use of nosological groups as an aid to the validation of questionnaires, or other testing devices, as long as the first deduction can be verified.

In their paper, Sigal et al. do not make explicit the important distinction between these two deductions, and in their discussion constantly switch from one to the other. When we look at their data, it is clear that these are confirmatory of the first deduction; as their Table 1 shows, dysthymics have the lowest extraversion scores (10.50), hysterics have higher scores (12.07), and psychopaths have the highest (14.88). In so far as this crucial deduction goes, therefore, the data do not seem to justify their conclusion that "the results suggest that either the hysterics and dysthymics cannot be used in the proposed manner, or that the E and N scales do not measure introversion-extraversion and neuroticism, or that both statements are true." As far as the first deduction goes the results suggest, on the contrary, that the extraversion scale puts the groups in the respective positions allocated to them by Jung's hypothesis, to wit, the hysterics and psychopaths towards the extraverted end of the scale, the dysthymics towards the introverted end. Needless to say, all neurotic groups have higher neuroticism scores than the normals—16.33 as against 11.62 for the combined groups.

The second error in their argument is related to what would presumably be their answer to the previous point. They would object, I imagine, that not all the differences I mention are in fact significant; indeed, only the dysthymic-psychopathic difference is statistically significant between the nosological groups. The answer to this point is relatively simple. A proper test of an hypothesis can only be made when *the number of cases used is sufficient to bring out the significance of the findings*. The number of cases used in this study is very small indeed; the total number is 52, and the number of psychopaths, to take but one instance, is only 8. When it is realized that the reliability of nosological classification is relatively low, and that consequently the criterion itself, even if valid, would not be highly reliable, then the need for much larger numbers becomes apparent. Sigal et al. might argue that their method of diagnosis was superior to that used by experienced psychiatrists of good standing, but such an assertion would require proof which is not contained in their paper.[2] If their findings had been directly contrary to the hypothesis, i.e., if the dysthymics had emerged as the most extraverted group, this lack of numbers might not have mattered so much. As it is, however, *their findings agree with prediction but fail to do so significantly*. One would have imagined that, before coming to such a strongly adverse conclusion, investigators concerned with the facts of the situation would have tried to obtain a larger number of cases to put the issue beyond a doubt.

This argument would be strengthened if it could be shown that other investigators were able to obtain similar results to theirs, and that the added cases rendered the observed differences significant. Fortunately, such further (unpublished) studies have been carried out in this department by A. Jensen, G. Claridge, and others, so that we now have available results on altogether some 200 neurotics, including the group tested by Sigal et al.[3] The total results, whether including their cases or not, are rather similar to theirs in placing the dysthymics at the introverted end, the psychopaths at the extreme extraverted end, and the hysterics in between; all differences between groups are statistically significant. The only apparent difference between these more inclusive results and those given by Sigal et al. lies in the fact that hysterics tested by them are lower in neuroticism than those tested later; this discrepancy is presumably related to their criteria of selection, which are a little unusual, and likely to lead to the exclusion of hysterics high on neuroticism. It may be noted parenthetically that when writing their paper, Sigal et al. were fully aware of some of these data, namely, those collected by A. Jensen, so that they must have been aware of the likelihood that if they had amassed data on larger numbers, their stress on the lack of significance in their data could not have been maintained.

We now come to the third point which will be dealt with briefly because it is still the subject of investigation. This relates to the fact that normal groups score at about the same level as do hysterics as far as the extraversion scale is concerned. Here, again, it may first of all be noted that the group of normals used by Sigal et al. is made up of students, and consequently not particularly well matched with the neurotic groups with respect to education, age, and class. It would have been better to have used the original standardization group, which was not only larger, but also much better matched for these variables (1). The student group used by Sigal et al., is slightly more extraverted and considerably more neurotic than the original standardization group; this fact may be in part responsible for some of the failures of Sigal et al. to obtain significant differences as, for instance, between the normal and the hysteric groups. They make no mention of this difference between their comparison group and the original standardization group.

However, when all is said and done, the more recent data on neurotic groups, as well as on much larger random samples of normals, leave no doubt that essentially Sigal et al. are correct in stating that *hysterics are not significantly more extraverted than normals* but that psychopaths are; dysthymics, of course, are significantly more introverted. There are two possibilities here. The first is that Jung was mistaken in his hypothesis about the position of normals, and the questionnaire data may be accepted as they stand. This view would in no way reduce the value of nosological groups for the purpose of validation studies, but would require a change in the statement of the theory regarding the degree of extraversion of hysterics as compared with normals. The second possibility is that there are certain distorting factors which make the hypothesis of *linear regressions* unlikely. There is some evidence to support the second possibility. In normal groups the correlation between extraversion and neuroticism is quite insignificant and in the neighborhood of −.1. In neurotic groups we

---

[1] Formally, the argument may be put in the form of an equation: extraversion:introversion = hysteria/psychopathy:dysthymia. It will be seen that the position of the normal group does not affect this formula, as this group does not enter into it.

[2] The unwary reader might assume that the diagnostic criterion used by Sigal et al. ensured high reliability because of the unanimity of three judges. This, unfortunately, is not so. It must be borne in mind that judges were not concerned with the patients and their symptoms in coming to a decision, but merely with the written records concerning each patient. However reliable the interpretation of the record may have been, this does not tell us anything about the reliability of the psychiatrist in arriving at his decision as to what to put into the record. The actual reliability of the diagnostic criterion would appear to depend much more on the latter than on the former.

[3] These data are summarized in the manual of the MPI, published by University of London Press, London, 1958. Also included there are data on criminals, reference to which is made below, and on psychosomatic patients; the large standardization group is constituted by a random sample of the population obtained by quota sampling methods.

have universally found the correlation to be significant and negative; the *r* values are usually in the neighborhood of −.3 to −.4. This result might be thought to be due to a selection factor that prevents subjects with high extraversion and high neuroticism scores from being sent to mental hospitals; they might be sent to prisons, for instance. However, recent large-scale studies of prison populations, carried out with the MPI, do not bear this out; scores of these groups are almost identical with scores of hysterics. Furthermore, it has been found that in normal groups the correlation between E and N in the 20% highest on neuroticism is again in the range of −.3 to −.4, although over the whole range it is between zero and −.1. Thus, clearly, one of the regression lines is nonlinear, and it is this failure of linearity which may in part be responsible for the failure of the hysterics to have more extraverted scores. It would be premature to speculate about the reasons for this nonlinearity; it may be due to such factors as response set (if hysterics have a response set towards answering "yes," then they would achieve high scores on neuroticism, but only average scores on extraversion where some "yes" answers are scored one way, some another), or it may be due to positive feedback in the reaction between strong innate emotionality and conditioning processes theoretically underlying the manifestations of dysthymia (2). Research is in progress at the moment to decide this point. At the present moment, it is certainly too early to come to any definitive conclusions on this point.

REFERENCES

1. Eysenck, H. J. The questionnaire measurement of neuroticism and extraversion. *Riv. Psicol.*, 1956, **50**, 113–140.
2. Eysenck, H. J. *Dynamics of anxiety and hysteria.* London: Routledge & Kegan Paul, 1957.
3. Sigal, J. J., Star, K. H., & Franks, C. M. Hysterics and dysthymics as criterion groups in the study of introversion-extraversion. *J. abnorm. soc. Psychol.*, 1958, **57**, 143–148.

Received June 9, 1958.

NEGATIVE EVIDENCE CONCERNING THE GENERALITY OF RIGIDITY[1]

DAVID R. FINK, Jr.[2]

University of Pennsylvania

For several years there has been lively debate on the question of whether behavioral rigidity, as measured by the familiar *Einstellung* test, operates as a general personality trait (4, 6, 7). The present study was undertaken to test for a general factor of rigidity among junior high school pupils.

Twelve experimental tests were developed on the *Einstellung* principle, in which a mental set was induced by a series of problems solved in the same manner (set problems), followed by a similar-appearing problem that demanded a new solution method (extinction problem). The test materials involved three types of reasoning ability—spatial, numerical, and verbal.

Two forms of each of the following instruments were devised: 1. Measures of spatial relations: Blocks Test, involving the Wechsler-Bellevue blocks; Figures Test, adapted from that reported by Schroder and Rotter (8); and Squares Test, adapted from Tate (9). 2. Measures of numerical reasoning: Number Series Test, and a modification of the Luchins Water Jar Test. 3. Measure of verbal reasoning: Disarranged Words Test. In the spatial tests, *S*s were required to complete a series of drawings or block designs of similar pattern, and then an extinction item demanding a shift of design or perspective. The Number Series Test required the completion of several series using the same solution formula, after which the solution method changed. Luchins' Water Jar Test was simplified to eliminate large numbers and the "critical" items. The Verbal Test presented a series of jumbled words which *S*s could unscramble in just one fashion, followed by an extinction word which demanded a new type of rearrangement. A fuller description of these measures is available in (3).

Rigidity effect on each test was measured by the time which *S* required to solve the extinction items. At least two types of scores can be taken from the raw times: (*a*) Extinction times alone may be considered; or (*b*) *S*s' time lag relative to set-item times may be computed by finding the

---

[1] This report is taken from a dissertation submitted to the Graduate School of Arts and Sciences, University of Pennsylvania, in partial fulfillment of the requirements for the degree of doctor of philosophy in education (3). The author wishes to state his indebtedness to M. W. Tate for providing constructive guidance throughout the study.

[2] Now at the University of Maine, College of Education.

TABLE 1

RELIABILITY COEFFICIENTS[a] OF RIGIDITY RATIOS[b] FOR EACH PAIR OF TESTS

Blocks A Blocks B	Figures A Figures B	Squares A Squares B	No. Series A No. Series B	Water Jar A Water Jar B	Words A Words B
.18	−.24	.56	.28	.08	.19

[a] Split-half coefficients, corrected by $\frac{2\,(r)}{1 + r.}$

[b] Median set-item time divided by extinction time.

ratio of median set-item time divided by extinction-item time (Rigidity Ratio).

The study was conducted in two stages, with samples drawn from the boys in Grade 8 at two suburban Philadelphia junior high schools. Only boys were selected in order to control any possible sex differences in the factor under consideration. Further, the IQ range was limited to five points above and below an assumed grade mean of 110.

It was hypothesized that if problem-solving behavior is affected by a unitary trait of rigidity, *S*s would tend to maintain their rank order time positions in solving extinction problems on the three types of tests.

EXPERIMENT I

Fifty-one boys were selected from the established IQ range. Mean age of the group was 13.4 years, with a standard deviation of 3.87. Mean IQ was 109.88, with a standard deviation of 3.93. Testing was conducted during October and November, 1956.

Although it was realized that the six pairs of tests were not in all cases true parallel forms, the instruments were thought to be similar enough to allow a test for consistency of rigidity effect in each of the three mental functions. The way would then be clear for an examination of rigidity consistency in all of the mental functions.

As shown in Table 1, the parallel-form reliability coefficients computed from Rigidity Ratios on each pair of tests provide no evidence of a general rigidity factor.

Despite certain test shortcomings, it was decided to carry out the proposed analyses in respect to the over-all consistency of rigidity effect. Owing to the nonnormal distribution of item times, it was appropriate to employ ranking methods. *S*s' time ranks were arrayed and analysis made of the differences among *S*s' mean or total time ranks. With the effects of the tests held constant, significant differences among the ranks would be indication of *S*s' holding their relative rank positions in all the tests. Such consistency would be evidence of a general rigidity factor.

Using the Rigidity Ratio (median set-item time divided by extinction time) as the criterion score, a rank test was run on the mean Ratios from each pair of parallel forms. All 12 test forms were included. The resulting chi square was 51.90, $n = 50$, giving $P = .41$. The differences were not significant.

Because of the marked differences in split-half reliabilities among the various rigidity tests, it was decided to analyze again after eliminating some of the low-reliability measures. Only one of these analyses produced significant differences. Raw extinction times for the five most reliable tests gave a chi square of 70.00, $n = 50$, with $P = .03$. This difference (significant at the conventional .05 level) among ranked extinction times encouraged speculation that a new sample and revised tests might produce more definitive results.

EXPERIMENT II

In the second experiment, the parallel-form approach was abandoned in favor of six revised single tests. Items observed to be ambiguous or excessively difficult were replaced. Set items were added, where necessary, so that each test contained six set problems. The new battery now consisted of improved forms of Figures Test B, Water Jar Test B, Squares Test B, Blocks Test A, Number Series Test B, and Disarranged Words Test A.

Thirty-one eighth-grade boys from a suburban junior high school similar to the original participated in January, 1957. The IQ range here was 101–119, expanded slightly over the first sample in order to get N above 20. Mean age for Sample 2 was 13.9 years, with a standard deviation of 6.32. Mean IQ was 110.16, with a standard deviation of 4.79.

Rigidity ratios were ranked for each test and analysis made of the differences between *S*s' total ranks. As in Experiment I, significant differences in these totals would be evidence of time consistency. The resulting chi square value was 16.92, $n = 30$, and $.975 > P > .950$. Again there was no evidence of the operation of a general rigidity factor.

SUMMARY

The question of whether rigidity effects tend to be generalized throughout most aspects of the individual's mental behavior remains confused. The present study can only be said to have added to the negative side of the argument. These data simply reinforce the position of Luchins (5) and others (1, 2) that rigidity is too complex and poorly defined at this stage to encourage any ex-

pectation of conclusive evidence concerning consistency of effects.

Perhaps the strongest features of the experimental design reported here are the close controls inherent in individual testing and the use of very similar tasks. If evidence of a general trait is not observable in these procedures, it is difficult to maintain expectations of discovering such evidence elsewhere.

REFERENCES

1. Applezweig, D. G. Some determinants of behavioral rigidity. *J. abnorm. soc. Psychol.*, 1954, **49**, 224–228.
2. Brown, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 469–476.
3. Fink, D. R. A study of the nature of behavioral rigidity. Unpublished doctoral dissertation, Univer. of Pennsylvania, 1957.
4. Luchins, A. S. Rigidity and ethnocentrism: A critique. *J. Pers.*, 1949, **17**, 449–466.
5. Luchins, A. S. On recent usages of the Einstellung-effect as a test of rigidity. *J. consult. Psychol.*, 1951, **15**, 89–94.
6. Rokeach, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.
7. Rokeach, M. Rigidity and ethnocentrism: A rejoinder. *J. Pers.*, 1949, **17**, 467–474.
8. Schroder, H., & Rotter, J. B. Rigidity as learned behavior. *J. exp. Psychol.*, 1952, **44**, 141–150.
9. Tate, M. W. Individual differences in speed of response in mental test materials of varying degrees of difficulty. Unpublished doctoral dissertation, Harvard Univer., 1947.

Received September 11, 1957.

COGNITION WITHOUT AWARENESS: SUBLIMINAL INFLUENCES UPON CONSCIOUS THOUGHT[1]

GEORGE S. KLEIN, DONALD P. SPENCE, ROBERT R. HOLT, AND SUSANNAH GOUREVITCH

Research Center for Mental Health, New York University

EVIDENCES of cognitive activity without awareness have often been reported, both in the clinic and in experiment. Systematic interest in this phenomenon was first aroused by clinical investigations of dissociated states and posthypnotic suggestibility (2, 20, 21). From the beginning, a cardinal proposition of psychoanalysis has been that the greater part of mental life goes on outside awareness. In his detailed discussion of the way dreams are constructed unconsciously from fragments of recent experiences—usually indifferent ones, Freud (10) cited examples in which the "day residue" had been a fleeting glimpse of something that had hardly entered consciousness. Poetzl (19) first demonstrated experimentally that the manifest content of dreams could reproduce stimuli that had never been conscious. After showing his subjects (*S*s) landscapes tachistoscopically for about 1/100 of a second, he had them describe and draw what they had seen. What was *not* reported showed up the next day in *S*s' accounts and drawings of their dreams. Poetzl's investigations have been repeated in recent years by Fisher (8, 9) and by Luborsky and Shevrin (22) with substantially similar results.

Another line of experimentation in the last twenty years has concentrated on demonstrating that object discrimination is not contingent upon awareness (1, 11, 14, 15, 17). In more recent studies, interest in nonconscious intake of stimulus properties has centered around the special question of defensive *reactions* to subliminal registrations, as in alleged instances of "perceptual defense" (3, 6, 15, 16). In the main, however, experimental studies have not carried inquiry about nonconscious registrations to a next logical step: how and in what ways do subliminal registrations affect conscious thought and within what limits?

The issue of "subliminal reception" and thought is highlighted by another suggestion that may be derived from Poetzl's results: that incidental, peripheral, or subliminal registrations may have unique effects upon thought; that absence of awareness may actually promote certain forms of participation of percepts in ongoing thought. Poetzl believed that subliminal registrations are more likely to serve as vehicles for dream thoughts than are conscious percepts.

A serious barrier to studies of these issues has been one of method. In the technique of which we have heard the most in recent years, used in studies of "perceptual defense," *S* is confronted with supposedly critical stimuli which are gradually raised from a low level of articulation to clearly recognizable levels. In this method, the nonconscious registration and the defensive reaction to it are *both* inferred from deviant recognition times. The threshold technique of these studies provides little if any information concerning the processes that determine the threshold measure. A new technique seemed necessary before nonconscious perceptual and cognitive processes could be studied experimentally in more detail.

To investigate these problems, we developed a method that records the impact of a subliminal registration upon the nature, extent, and quality of an *S*'s conscious impressions of a perceived stimulus. The technique does not require *S* to report directly on the critical subliminal stimulus; rather, the influence of subliminally presented figures is determined by the study of associated changes in *S*'s responses to a consciously seen figure.

---

[1] An abridged account of this study was presented at the Sixty-Third Annual Convention of the American Psychological Association, 1955.

The study is one of a series supported by a grant from the National Institute of Mental Health (Project Code Number: M-782, Cognitive Processes Below the Level of Awareness), and was carried out in the Psychosomatic Laboratory and Department of Psychiatry, School of Medicine, State University of New York. We are grateful to Robert M. Dickes, who was then Director of the Psychosomatic Laboratory, for generously making available the Laboratory's facilities and giving us access to the student body of the medical school for experimental *S*s, and to Michael Tobin and others of the Laboratory staff for their freely offered assistance at various points in the study. Our appreciation also goes to Jan Frank of the Department of Psychiatry for useful exchanges of views regarding interpretations of the data.

We wish to thank Sheldon Bach, Reeva Safrin, and Joan Havel who rated the protocols; and Bessie Boris, Worden Day, and Leo Garel, the artists who drew the stimuli and helped develop the check list for evaluating the *S*s' drawings. We are also indebted to Harriet Linton and Gudmund Smith for their assistance as judges.

The technique is based upon a phenomenon described by Werner (25) and later by Cheatham (5). Expose in quick succession a pair of figures at the same point on the screen of a mirror tachistoscope. If the two follow each other in sufficiently rapid succession and if the second appears in a slightly longer exposure, only the second of the pair (*B-figure*) is consciously seen. Complete "obliteration" of the first figure (*A-figure*) occurs, even when it is exposed so long that it could be clearly recognized if it were not succeeded by the second figure.

Werner and Cheatham considered relative contour strength and spatial disparity of the critical A- and B-figures to be critical variables in accounting for the phenomenon. But if the phenomenon can be obtained with complex figures when such congruities of contour do not exist, the way is opened to use it as a means of studying issues of "subliminal registration." Pilot studies strongly suggested that the A-figure, even though not seen, exerted an effect upon conceptions of the B-figure. If so, it would be possible to vary systematically the *meaning* of the masked A-figure and to look for distinctive effects of such meanings upon impressions of the consciously perceived second figure.

The present study centered upon two main questions: (*a*) Will impressions of a pictured person (B-figure) of ambiguous sex be significantly affected by subliminally presented sexual drawings? (*b*) Will impressions of the B-figure be differently affected by realistic and by symbolic sexual pictures shown as A-figures? When sexual connotations of a form are presented symbolically and thus disguised, its effects upon thought may be distinctively different from those of uncamouflaged, realistic forms.

METHOD

Apparatus

An electronically controlled mirror (Dodge) tachistoscope was used which consists of three fields: a blank field on which *S* is ordinarily focussed, and two stimulus fields.[2] The rest of the interior is painted flat black or covered with black velvet. The blank field consists of a white cardboard on which is drawn a black rectangle circumscribing the area where the stimulus figures appear.

Each stimulus compartment is equipped with two mercury argon tubes surfaced with a special phosphor coating that produces a daylight white color, with rapid onset and quick decay so that the light output closely approximates a square wave form even at very brief exposures.[3]

The illumination of each field can be independently varied. All fields were adjusted to equivalent brightness by means of a photometer placed at the viewing aperture. In the present study, the brightness values of the three fields were fixed at .8 apparent foot candles for the range of exposure times used (as determined by a Macbeth Illuminometer).

The electronic timing unit fires the light source in the three fields independently in predetermined sequences. Durations of the two stimulus fields can be varied continuously from .002 to 5.9 seconds. The accuracy of the time intervals is ±1 per cent above .10 sec. and ±2 per cent below .10 sec. The two stimulus fields may be exposed singly or in succession with no intervening intervals. The blank field reappears after each stimulus or stimulus sequence. The start of the sequence is controlled manually.

Each stimulus compartment accommodates 7″ square stimulus cards. The stimulus fields are arranged in the box at a 90° angle from each other at an optical distance of 3.5′ from the eyepiece. Coated low-reflection mirrors with a transmission-reflection ratio of approximately 50-50 are carefully fixed in the box so as to transmit the two stimuli over a common optical path to *S*'s position. The placement is such as to equate the retinal sizes of each of the stimulus fields.[4]

Design and Procedure

The B-stimuli, shown in Fig. 1, were line drawings of three human figures whose sexual identity was intended to be uncertain. The A-stimuli were India ink line drawings of male genitals; female genitals; a "masculine symbol" (trylon and perisphere); and a "feminine symbol" (a shield-like triangular form). The A-figures are shown in Fig. 2. The A-stimulus, whether realistic (male or female genitals) or symbolic, was always placed in the center of the field. Thus, the genitals were anatomically displaced with respect to the B-figure.

The symbols were chosen from among a larger number of possibilities after we found that pretest *S*s,

---

[2] A complete description of the tachistoscope and electronic timing units is obtainable from Fred and Charles Snyder, The Menninger Foundation, Topeka, Kansas, who designed and constructed them.

[3] We are grateful to Lawrence Karlin for his help in finding white-light tubes with desirable onset and decay attributes and for checking their wave forms. The tubes are described in Karlin (12).

[4] The adjustable card carriers make it possible to displace each of the two fields laterally when this is desired. The maximum lateral displacement of the two fields is one-half of a 7″ card.

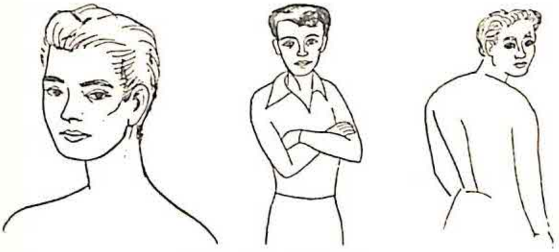

FIG. 1. B-STIMULI

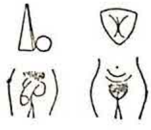

FIG. 2. A-STIMULI

asked to associate to them, gave consistently female sexual associations to the female sexual symbol and consistently male ones to the trylon and perisphere (though for the latter, obvious sexual meanings appeared only later in the flow of associations).

The four A-figures were paired with the three B-figures and shown subliminally to different *S*s and in different pairings in a balanced design shown in Table 1. A control condition was also included in which B-figures were shown without a preceding A-figure.

Each *S* was tested in two sessions. In Session I he was shown a B-figure with no preceding A-figure (control condition) and then another B-figure with a preceding subliminal A-figure (e.g., male genitals or female genitals; male symbol or female symbol). For half the *S*s the control-subliminal sequence was reversed. In Session II, which followed the first by three to ten days, *S* saw a mirror image[5] of the *same* B-figure with the opposite member of the first subliminal series, and then the third B-figure paired with a supraliminal A-figure.[6] The balanced design provided for equal distribution of all pairings and conditions among the group of 24 subjects. The conditions and numbers of subjects in each are summarized in Table 1. *S*s were assigned to conditions randomly. In the case of 17 *S*s it was possible to repeat the control (blank) condition in the second session as well. Following the test conditions in the second session, ascending recognition thresholds were determined for the A-figures presented *alone* with each *S* tested on stimuli that had been shown him as A-figures in the subliminal conditions.

B-figures were always exposed for .33 second, far above recognition threshold and providing *S* with a clear but brief look at the stimulus. The A-figures were exposed in the subliminal condition at much briefer exposures, varying slightly from *S* to *S* within a range of .02 to .035 second. The exposure time for subliminal A-figures was tailored to each *S*'s recognition threshold for geometric figures (a diamond and a cube) which was determined prior to the critical experimental trials in the first experimental session. To make allowances for discrepancies of form between the geometric figures and the critical A-figures, a constant, arbitrary value of .015 second was added to the higher of the *S*'s recognition threshold for the geometric figures. For instance, if *S*'s recognition threshold for the diamond was .005 second and for the cube .01 second, addition of the constant gave a value of .025 second, which was then fixed as his exposure time for all A-figures shown to him in the subliminal conditions. In all A-B pairings, the B-figure immediately succeeded the A-figure so that the only phenomenal change was the sudden brief appearance of the dark lines of the B-figure in an otherwise constant field. Between the exposure of stimulus pairs, the blank field remained on steadily. General illumination in the experimental room was low.

The first experimental session began with the measurement of *S*'s ascending recognition threshold for the

TABLE 1

EXPERIMENTAL DESIGN

(N = 4 for each of six conditions)

Session 1		Session 2	
Subliminal[a]	Control	Subliminal	Supraliminal[b]
M gen-I	II alone	F gen-I	M/F sym-III
M gen-II	III alone	F gen-II	M/F sym-I
M gen-III	I alone	F gen-III	M/F sym-II
M sym-I	II alone	F sym-I	M/F gen-III
M sym-II	III alone	F sym-II	M/F gen-I
M sym-III	I alone	F sym-III	M/F gen-II

M gen = male genital A-stimulus
F gen = female genital A-stimulus
M sym = male symbolic A-stimulus
F sym = female symbolic A-stimulus
M/F = male or female
I = B-picture (face)
II = B-picture II (figure with arms folded)
III = B-picture III (figure with back turned)

[a] Half the *S*s had a male A-stimulus in Session 1, half a female A-stimulus. When male was presented in Session 1, female was presented in Session 2 and vice versa. Also, half the *S*s in each sequence had the subliminal presentation first in this session, half had the control first.

[b] The A-stimulus presented to seven *S*s deviated from the general design: they were given one of the A-stimuli which they had been shown earlier under subliminal conditions.

---

[5] The repeated B-figure was shown in mirror-image in order to make any perceptual differences caused by the change of the A-figure more plausible to the *S*. Care was taken to balance all variables so that there was no confounding of sequence (M to F or F to M), type of A-figure, and orientation of B-figure. Since no systematic differences in the ratings of drawings of the B-figures were associated with their being in the original or the mirrored orientation, no further reference will be made to this factor.

[6] Results with the supraliminal pairing of A- and B-figures are not included in the present report because of an unanticipated flaw in the design. It turned out that reactions to the different B-figures differed so much that all comparisons we could make between effects of subliminal and supraliminal pairings were confounded with B-figure differences. Separate analyses for the different B-figures were precluded by the small number of *S*s in each stimulus in the supraliminal condition. We plan to compare supraliminal and subliminal effects more intensively in another study.

cube and the diamond. *S* was comfortably seated so that he could rest his forehead against the eyepiece and look with both eyes (keeping his glasses on if he wore them) through the black tunnel of the tachistoscope to the moderately bright, blank fixation field. After the subliminal A-stimulus exposure time was found in the manner described above, *S* was next told:

> You are going to see a figure that will appear briefly. I will show it 10 times and each time I will say "Ready," just before I show it to you. I want you to look at it carefully because later on I will ask you to describe the personality of the figure, the type of person the figure seems to be. I'll also ask you to draw it.

There were then 11 exposures (one to prepare *S* plus 10 more) at 2-sec. intervals of either a subliminal A-figure followed by a B-figure, or a B-figure alone (control condition). *S* then went to an anteroom where he filled out rating forms describing the "personality" of the B-figure, made a free drawing of it, and was interviewed on other aspects of his impression.

In the second session, *S* was shown a new A-figure in a subliminal pairing with the same instructions. He was also shown a new pair of A- and B-figures in a supraliminal presentation.

To review briefly the test conditions for an individual *S*: *Session I:* recognition thresholds were determined for geometric figures. Then followed the subliminal A-B pairing and the control B-figure (or vice versa). *Session II:* subliminal A-B pairing, followed by the supraliminal A-B pairing and the control B-figure again. Finally, recognition thresholds were obtained for each of the A-figures that had been shown to *S* in the subliminal condition.

Subjects

*S*s were 24 male medical students in the third year of training. During the period of the experiment, they were engaged in a one-month psychiatric "clerkship" and their participation was made part of the course requirement. They were told only that it was an experiment in perception.

RESULTS

Principal Response Variables

Our major hypothesis was that impressions of the B-figures would incorporate expressive qualities and meanings of the subliminal A-figures. Our first, simple hypothesis was that the subliminal figure would cause *S* to see the B-figure as having the corresponding sex. Since all *S*s tended to see all B-figures as males, we had to look for effects in subtler measures than an indication of sex. Thus, we expected that a B-figure, presented contiguously with the male genital, would acquire more attributes that were distinctively associated with the male genital than it would when paired with the female genital. Essentially the same trends were expected with the symbols.

The *S* recorded his impression of the B-figures in two ways: verbal descriptions (a personality check list) and drawings of the B-figures.

Personality check list. This check list consisted of 69 pairs of adjectives, e.g., "modest-boastful," each pair forming the extreme positions of a 4-point scale. Items were chosen that seemed applicable to all three B-figures and at the same time might be sensitive to the effects of the subliminal A-figure. Osgood's list (18) was a principal source of items.

The same personality check list was employed in developing scoring keys on the expressive qualities of the A-stimuli. Twelve male judges, graduate students in clinical psychology, described all four A-figures. We pooled the description of each panel, disregarding differences between the "somewhat" and "very" ratings, and compiled into clusters those items that at least 10 out of 12 judges applied to the male genital, the female genital, the male symbol, and female symbol. The clusters were further paired by eliminating items common to each pair, with the result that the final clusters represented distinctive, nonoverlapping attributes which the given stimulus provoked in the majority judgment. The four A-figure keys were used in the analysis of *S*s' descriptions of B-figures on the check list and are presented in Table 2.

Drawings. In the subliminal conditions, each *S* made two drawings of the B-figure, one in the male A-figure condition, one in the female A-figure condition. Each pair of drawings was compared in *manliness* by five psychologists who made a "more than"–"less than" judgment. The number of judges making the discrimination in a particular pair of drawings was used to weight the difference; the more judges agreeing on a pair, the higher the relative masculinity score. Thus, highest rank went to an *S* whose drawing in the male genital condition was judged more manly by all five judges. Lower ranks were given to *S*s whose drawings in this condition were judged more manly than the other by less than unanimous decision (i.e., 4-0, 4-1, 3-0, 3-1, 2-0, 2-1). Male genital drawings judged *less* manly were ranked in reverse order in the same fashion. In this way, the 12 pairs of drawings could be ranked for comparative masculinity in each condition. (We will refer to this measure

TABLE 2

A-STIMULUS KEYS

(Unique characteristics of the A-figures as judged by 10 out of 12 judges)

Male Genital	Female Genital	Male Symbol	Female Symbol
boastful	womanly	brave	womanly
brave	peaceful	clear-thinking	loving
happy	loving	manly	subtle
manly	contented	strong	warm
strong	delicate	ambitious	kind
ambitious	soft (opposite loud)	decisive	rounded
decisive	intuitive	active	generous
active	smooth	rugged	
rugged	sweet	clean	
outgoing	soft (opposite hard)	calm	
dominant	vivacious	outgoing	
hard	kind	dominant	
large	rounded	vigorous	
optimistic	generous	independent	
mature		hard	
		large	
		effective	
		angular	
		mature	

as Comparative Masculinity.) This method of evaluating the drawings assumed that the judges' conception of manliness would be related to the effects of the A-stimuli. A drawback of this score was that it was based on a comparison of *S*'s own drawings only. One cannot tell how *S*s compared with each other in respect to the manliness of their drawings.

Therefore, another technique of assessing manliness was used. Drawings of B-figures made by *all* *S*s in each subliminal condition were ranked in respect to manliness. (We will refer to the measure as Relative Masculinity.) Two judges made each ranking and their combined ranks was the final score for each *S* on a given condition.

Subliminal Effects: Realistic A-Stimuli

Before we go into the effects of the subliminal stimuli, consider for a moment *S*'s task: he was given a short series of glimpses of a line drawing; a few minutes later he had to make a rather detailed set of ratings and then draw his fast-fading impression of what he had seen. We expected that some attributes of the A-figure would be incorporated in *S*'s check-list description of the B-figure. Therefore, when a B-figure was paired with the male genital, there should be more items from the male genital key in descriptions of the B-figure then when the same B-figure is paired with the female genital.[7] An agreement is scored if *S*'s reaction to the B-figure on any item agrees with the key. The degree to which the male stimulus was *uniquely* effective in provoking the qualities associated with it in the key was found by subtracting agreements with the male key in the female stimulus condition from the number of agreements in the male condition; these differences were ranked for the 12 *S*s.

Agreements with the A-figure key were higher in the male A-stimulus condition for six *S*s, lower for three *S*s, and unchanged for three *S*s. Assuming that the key is a reasonable summary of connotations provoked by the A-stimulus, it is evident that not all *S*s incorporated these qualities in their descriptions of the B-figure. *S*s were *consistent*, however, in their tendency to do so or not. It is interesting that all the reversals of this tendency were consistent, i.e., if *S*s responded to the male A-figure with qualities of the female genitals, they responded to the female A-figure with qualities of the male genital. Hence, it is not surprising that the degree to which *S*s brought qualities of the stimulus key into their descriptions of B-figures is correlated .82 in the male and female genital A-stimulus conditions. If the male genital induced *S* to incorporate male genital qualities, the female genital was likely to be equally effective in coloring impressions of the B-figure with its associative qualities.

Further evidence for a consistent reaction of the *S*s in the two subliminal presentations of the genital stimuli comes from the drawings. As the correlations in Table 3 attest, *S*s who attributed characteristics of the subliminal male genital A-figure to the paired B-figure also produced the most manly drawings in response to the male A-stimulus (i.e., they ranked highest on Comparative Masculinity). This appropriate differentiation of drawings according to the A-figure exposed also correlated significantly with the check-list measure of the effectiveness of the female genital. Thus, *S*s differed considerably in their sensitivity to influence by the subliminal genital stimulus,

---

[7] Comparisons of the subliminal conditions with the blank condition were difficult to make because the B-figure paired with the blank was not the one presented to *S* in the subliminal conditions. Since different B-figures provoked varying impressions on the check list, it was difficult to assess the unique effects of subliminal stimuli in such a condition.

TABLE 3

RANK-ORDER CORRELATIONS BETWEEN MAJOR VARIABLES

Realistic Stimuli	
Male genital effect—Check list vs. Female genital effect—Check list	.82***
Male genital effect—Check list vs. Comparative masculinity—Drawing	.64**
Female genital effect—Check list vs. Comparative femininity—Drawing	.54*
Male genital effect—Check list vs. Relative masculinity—Male genital drawing	.51*
Female genital effect—Check list vs. Relative femininity—female genital drawing	.18
Symbolic Stimuli	
Male symbol effect—Check list vs. Female symbol effect—Check list	.34
Male symbol effect—Check list vs. Comparative masculinity—Drawing	−.12
Female symbol effect—Check list vs. Comparative femininity—Drawing	−.64**
Male symbol effect—Check list vs. Relative masculinity—Male symbol drawing	.24
Female symbol effect—Check list vs. Relative femininity—Female symbol drawing	−.27

* $p < .10$.

** $p < .05$.

*** $p < .02$.

Note.—For 12 cases, a rho of .51 is significant at the .10 level, a rho of .57 is significant at the .05 level, and a rho of .71 is significant at the .02 level, all two-tailed tests.

but they tended to be consistent in the amount of effect they showed in both graphic and judgmental behavior (drawings and check-list ratings).

We can also look for consistencies of reaction within a single session rather than in comparisons of sessions, as in the data just cited. The Relative Manliness of the drawing was correlated with the check-list measure of the effectiveness of the male genital A-figure at a near-significant level, though this same consistency did not show up in the session when the female genital was the A-figure (Table 3). Evidently the manliness induced by the subliminal male genital is the effective agent in producing the differences between the two *S*s' drawings with respect to masculinity. Because the B-figure is not truly ambiguous but usually seen as a man (so described by eight out of twelve *S*s in the blank condition) it may be more susceptible to the effect of a male A-stimulus than to a feminine A-stimulus. Of course, the male genital may be generally a more powerful subliminal stimulus for this sample of subjects; however, this seems unlikely in view of the high correlation (.82) between sensitivity scores on the check list under the two conditions. At any rate, the subliminal female genital did not produce the expected increase in the femininity of the B-figure *drawing*.

Self-image. One source of variance is the masculinity of *S*'s self-image. Several weeks after the experiment, *S*s were asked to describe themselves on the same personality check list used in the main experiment. Using the two genital keys as extremes of masculinity and femininity, we tallied agreements with *S*'s ratings of self on the same check list. *S*s with relatively *feminine* self-ratings made very masculine drawings of the B-figure following the subliminal female genital. Relative Masculinity of drawings in this condition correlated .62 with relative femininity of self-image ($p < .10$, two-tailed test. All significance levels below are two-tailed). Conversely, a relatively masculine self-image is associated with femininity of the drawings (same correlation). In other words, the female genital provoked the expected feminine drawings in those *S*s whose masculine self-picture was fairly clear, whereas it produced a more masculine drawing among *S*s with a relatively large feminine component in their self-conception.

Subliminal Effects: Symbolic A-stimuli

In order to compare effects of the two main types of A-figures, analysis of the symbols followed the same lines described for the genitals, and starts with the two stimulus keys. As Table 2 shows, the male symbols represented largely masculine qualities to our panel of judges, supporting our assumption that the trylon and perisphere is also a male symbol. To a much lesser degree, the so-called feminine symbol was assigned feminine connotations. The relatively small number of items that *uniquely* described the female symbol key indicates the low degree of uniformity of meanings possessed by this abstract form.

Neither A-figure had a distinctive effect consistent across *S*s on check-list impressions of the B-figure; in this respect, reactions to symbols resembled those to realistic figures. Six *S*s incorporated more aspects of the subliminal male symbol in their B-descriptions under the male symbol condition; four incorporated fewer aspects, and two showed no change. Only four *S*s showed the expected effect in the female symbol condition. Also indicative of this inconsistency between stimulus reactions is that reversals of expectation in one condition did not always occur with reversals in the other condition. This contrasts with the high inter-S consistency on the genital stimuli. The correlation between the male and female symbols with respect to the distinctive effectiveness of the A-stimuli was not significant—only .34.

The drawings were more in accord with the hypothesis of a direct subliminal influence of the symbols. In nine of the twelve *S*s, drawings of the B-figure in the male symbol condition were more masculine than corresponding drawings under the female condition ($p < .04$, sign test two-tailed). The male symbol, sharing many qualities with the male genital according to the panel of 12 judges, appeared to influence the B-figure directly and to provoke a relatively more masculine drawing. Moreover, the effect was much more uniform than that of the male genital which produced more masculine drawings in only 7 out of 12 *S*s.

Comparative sexuality of the drawing tends to be negatively rather than positively related to responsiveness to symbol attributes on the check list. Comparative Masculinity of the B-figure drawing following the male symbol is correlated $-.12$ with agreement with the male symbol key; Comparative Femininity of the B-figure drawing following the female symbol is correlated $-.64$ with agreements with the female symbol key. Evidently, the subliminal female symbol provoked a relatively more feminine drawing of the B-figure in *S*s who described it as more masculine on the check list, suggesting that the effect of the subliminal stimulus was expressed in only one of these response channels for most *S*s.

Possibly the differences between results with check list and with drawings stem (*a*) from the distinctive linear qualities of the two symbols and (*b*) from the differences in covert and overt meanings of the two symbols. A distinctive feature of the male symbol is its long triangular form; a distinctive feature of the female symbol is its curving lines and general symmetry. The genital stimuli, by contrast, have many more figured similarities. These differences in linear quality of the symbols may have been especially evocative of different graphic responses, in ways that were relatively independent of the sexual meanings. On the other hand, the divergent meanings that *S*s assigned to the female symbol make for a less uniform, relatively unstereotyped (verbal) impression (see Table 2). The male symbol clearly had an obvious, social meaning (trylon and perisphere) which is certainly more accessible than its covert sexual connotations. The female symbol, not having such a clear socially stereotyped meaning, produced a considerable variety of interpretations and was hence susceptible to idiosyncratic interpretation.

The self-image data suggest that certain *S*s are more sensitive to the feminine aspects of the female symbol, namely *S*s with a highly masculine self-image. *S*s whose self-ratings contained a large number of items from the male genital key were highly sensitive to the female symbol; many of the distinctive attributes of the female symbol appeared in their descriptions of the associated B-figure. Correlation between masculinity of self-image and sensitivity to the female symbol is .75, $p < .02$. Evidently for certain *S*s, a clear sexual meaning in the female symbol is subliminally effective. However, for the symbol group as a whole, form seems to be a more clearly effective determinant of the response and tends to obscure the particular contribution of the sexual meaning.

Recognition Thresholds

We will now consider whether or not indices of threshold sensitivity to the A-figures were correlated with the behavioral effects that were central to the present study. Thresholds for both sets of A-figures were obtained at the end of the second experimental session, with threshold defined as that exposure level at which S could accurately identify the A-figure. The lowest exposure was 4 milliseconds and exposures were increased in an ascending series in logarithmic progression until recognition was achieved. On the genitals, we considered recognition to be complete when S said "penis" or "male organs" or "vagina" or "female organs"; on the symbols, when S was able to say "trylon and perisphere" or "circle and triangle" for the male symbol, and accurately describe the inner details of the female symbol. Order of presentation was random for each group of 12 *S*s, with 6 *S*s receiving the male symbol first and 6 the female first; 5 the male genital first and 7 the female genital first. For each *S*, thresholds were obtained only for the subliminal A-figures used in the main experiment.

We will consider first the relationships between the genital thresholds and check-list reactions to the subliminal male and female pairings. Threshold sensitivity was unrelated in any significant way to distinctive responsiveness to the subliminal stimulus. Sensitivity was, however, related to the degree to which *S*s responded differentially to the A-figures on the check list, regardless of whether their reactions conformed to the stereotype of the A-figure. Average threshold for the two genital A-figures was significantly correlated with this measure of differential responsiveness on the check list (rho = $-.61$, $p < .05$), with *S*s having high thresholds for both genital figures giving the least differentiated responses to the two genitals in the subliminal condition. Sensitivity to the male genital seems to be the major factor in this relationship, for male genital thresholds were significantly correlated with differential responsiveness (rho = $-.59$, $p < .05$) while female genital thresholds were not (rho = $-.24$).

Turning to the A-figure effects on the drawings, we find a more complex relationship with thresholds. High threshold alone is not related to graphic measures of subliminal sensitivity, but *S*s with either extremely high or extremely low thresholds to both genital stimuli drew noticeably more masculine figures in the male subliminal condition than did *S*s with more average thresholds ($p = .05$, Mann-Whitney test). The more deviant the threshold, the more the drawing reflects the subliminal figure.

Recognition thresholds for the symbols were related in a markedly different manner to subliminal effect scores. Whereas genital thresholds showed some relationship to check-list responses, threshold level for the symbols was related only to the drawings. Low threshold for the male symbol was correlated .52 ($p < .10$) with Relative Femininity of the drawing in the female symbol condition, and low threshold for the female symbol correlated a .60 ($p < .05$) with Relative Masculinity of the drawing in the male symbol condition. The direction of these relationships is not readily interpretable. We are able only to say that high threshold for one symbolic figure is related to sensitivity to the opposite symbolic A-figure in the subliminal condition.

Two general trends stand out. Recognition thresholds for the genitals and the symbols are each related to subliminal behavioral effects, but in different ways. Check-list responsiveness is generally related to thresholds for the genital figures while responsiveness on the drawings is related to thresholds for the symbols. It may be that the threshold for the genital figures is primarily a threshold for meaning and for this reason relates most directly to the check-list indices of subliminal effect. Symbolic thresholds are more clearly thresholds for form, and understandably are related more directly to the masculine and feminine properties of the drawings which were judged on a largely formal basis, using such categories as roundedness of line, symmetry, etc.

DISCUSSION

The main implication of the findings is that stimuli may acquire psychological representation without *S*'s awareness, and subsequently affect cognitive and behavioral processes to a measurable degree. There is little question, however, that the principal effects appeared only in subtle measures and were complicated by idiosyncratic interactions among the variables. Does such a seemingly unstable phenomenon merit further concern? In our opinion it does. Important theoretical issues are perhaps concealed by the fact that the *unconscious* influences on perceptions *are* small under ordinary circumstances. Were it possible now to maximize the effects by manipulating appropriate variables, such issues might be brought into relief.

We will consider some of the variables and issues that were especially highlighted by the present findings, and indicate some possibly fruitful directions for further study.

1. *Meaning and form in the effects of A-figures.* In their simplest guise, the results suggest that the A-figure interacts in some manner with the B-figure to produce a response that combines or in some way uses the elements of the two stimuli. Both meaning and form seem to be crucial determinants; the different effects of the two variables are highlighted in the results with realistic and symbolic A-figures.

The most obvious difference in the effects of the two types of A-figures was that the genital stimuli produce consistent results on both drawings and check list, whereas the symbols' effect is pronounced only on the drawings. In the symbol condition, the one sizable trend was that a significant majority of *S*s drew the male symbol B-figure as more masculine than the female symbol B-figure. Group effects on the drawing (and also the check list) were nonsignificant with the genital A-stimuli. However, individual differences were more meaningful, as evidenced by significant correlations between drawing and check list, and indicated that the A-stimuli were effective though in idiosyncratic ways.

The greater consistency of results for the genital condition appears to be related to the greater coherence of the genital A-stimuli, in the sense that the form of the genitals is inextricably tied to their meaning. Considered simply as stimuli, they are unambiguous; there is no escaping their crude sexual message. If we assume that formal effects are represented primarily on the drawings, and content effects on the check list, the high correlation between the two measures seems highly reasonable.

The symbols are quite a different set of stimuli. The spherical triangle (see Fig. 2) does not unambiguously represent femininity; a glance at the group consensus shown in Table 2 points up the lack of agreement in how the judges interpreted the female symbol. There is less *invariance* between meaning and form in the case of the symbols; they have other connotations besides sexual ones, and we have no guarantee that sexual connotations were primary. The forms might have sexual equivalents for some *S*s and not for others, and for this reason, correlation between drawing and check list would be low.

Because the connotations of the symbols are diverse, the forms of these stimuli are more independent as stimulus properties. Now it is possible to assume that purely formal aspects of the A-figure interact with formal aspects of the B-figure so that the response represents the direct effect of one contour on another. The more masculine drawings following the male symbols may represent such an interaction; the size of the effect may stem from the fact that because form is the main variable, individual associations to the symbols are not involved.

If the more uniform drawing results do indeed reflect the formal properties of the symbols, it would seem that a stimulus with pronounced form and disguised or ambiguous content has a definite subliminal advantage. Perhaps ambiguity of meaning helps to bypass individual associations which tend to complicate the group results. It is also possible that formal stimulus properties are less affected by the high exposure speed used in the tachistoscopic presentation, and preserve more of their original stimulus properties, whereas more meaningful stimuli may be differently affected by different speeds.

2. *Nature of subliminal interaction.* How, exactly, do the A and B processes interact? At least two possibilities are suggested by the results. In some *S*s, elements from A and B stimuli fuse to produce a condensation, as when the drawing of the ambiguous B-figure becomes more masculine when it follows a male symbol. In other *S*s, the A and B elements are kept separate, with the properties of the A-figure serving as a ground against which the B-figure is judged. In this case, the ambiguous B-figure is seen as more feminine following the male genital, and its difference from the sex of the A-figure is emphasized. As noted above, some *S*s included attributes of the genital figures in their impressions of the B-figures, others excluded such attributes, and others were not affected, with a high consistency in individual responses. If *S* included attributes of the male genital, he did the same with the female genital. These consistencies strongly suggest that inclination to condense or contrast the A and B stimuli may be a personality variable. Different responses to the subliminal genitals seem to depend partly on the feminine component of the self-image, and more generally, on *S*'s overall sexual adjustment. We know that *S*s with a well-established masculine self-image tend to incorporate elements of the subliminal female genital stimulus in their drawings of the associated B-figure, whereas *S*s with less masculine self-image tend to exclude feminine elements. Strength of the masculine self-image in male *S*s is certainly related to general sexual adjustment; a low masculine self-image hints at specific conflicts in this area. These

conflicts may be aroused by a subliminal female genital and lead *S* to deny the female aspects of the B-figure. We suggest the possibility, therefore, that the "choice" of condensation or contrast-formation as the way a subliminal stimulus interacts with conscious cognition may vary with the stimulus material, but that for stimuli like ours, it may be a function of defenses associated with self-identity.

Interactions of the A and B processes may be represented in different qualitative experiences of the B-stimulus. The interaction may be reflected in *S*'s perception of the B-figure, in judgments, in feelings, or in some combination of these. It may occasion a variety of verbal and motor reactions. Thus, the interactions between the two processes need not be represented only perceptually but may appear in various forms of behavior or experience.

The qualitatively different outcomes of the A and B interaction are communicated to *E* through different channels, some of which involve awareness, others not. It seems worthwhile to distinguish expressive channels of report from the experiential and nonexperiential qualities of *reaction* to B. In the present study for instance, it cannot be said to what extent the check-list responses or the drawings represent perceptual experience, judgments, or imaginal elaborations. Moreover, some forms of report convey certain qualities of the experience more easily than other forms of report. Description of a face in a set of fixed categories offers very limited means for conveying a variety of reactions as compared to drawings. This differential sensitivity may explain the inconsistency between the drawings, check list, and other forms of report, and suggests that knowledge of subliminal processes waits on the development of more sensitive measures.

3. *Recognition thresholds and behavioral effects.* Comparison of the recognition threshold data of the present study with changes in the B-figures points up some of the drawbacks of studies which have sought evidence of "defense" solely in recognition threshold level (e.g., 6, 15, 16). Our findings suggest that the recognition threshold is only a limited and for certain stimuli even an insensitive indicator of the personal significance of the stimulus. For instance, the symbols had rather clear effects both in the check list and in the drawings, but their thresholds were not especially elevated or depressed, and individual differences in thresholds were related to differences on the drawings alone and then only in ambiguous fashion. Even the more widely varying recognition thresholds for the genital stimuli did not reliably indicate their potency in other areas of response. As a whole, the results suggest that the modes of cognitive elaboration which a stimulus may undergo below awareness are only coarsely represented in the narrow datum of recognition level.

4. *Discrimination and registration.* We have assumed that in the masking conditions of the present study, exposure of the A-stimulus occurred at limits beyond *S*s' capacity to discriminate the figures, either in part or in whole. Our findings, like those of Fisher (8, 9) and Luborsky and Shevrin (22) support a conception of *registration* (reception, organization of intake, trace activation) without partial or complete perception.

On the other hand, there are those who maintain that it is unnecessary to assume two processes (4, 7, 23). Their argument runs something like this: Only a discriminable stimulus can affect thought, and discrimination always depends upon at least some information being within *S*'s conscious grasp. Therefore, if the A-stimulus has exerted any effect it could only be because just enough of its properties were partly or dimly accessible to make possible the discrimination required for the stimulus to have any further behavioral impact. Adams, in a recent review of studies that purport to demonstrate discrimination without awareness, concluded that ". . . the only type of behavior without awareness which can be easily reproduced on the basis of published reports is the classical type, in which *S* knows *what* he is supposed to be discriminating, but does not know *that* he is discriminating, because of the absence of the usual sensory experiences to which he is accustomed under the given type of stimulation" (1, p. 402). In the present study, *S* certainly did not know *what* he was supposed to be discriminating in the "subliminal" stimulus, nor even *that* he was responding to a subliminal stimulus, for the test conditions did not require him to respond *directly* to it (in contrast to the usual discrimination-without-awareness studies) but only to a contiguous, consciously seen form. In spite of this, we were able to observe changes in reported impressions of the consciously seen figure, and to relate these to qualities of the A-figures.

The crucial question that remains is whether *S was* partially aware of the A-stimulus but did not report it. The issue is clearly drawn. The assumption that conscious perceptibility is not a necessary condition for registration has considerable implications for a general theory of cognition. But before such an unusual assumption can be accepted, it is necessary to rule out the simpler explanation that *S* has some partial awareness of the so-called subliminal stimulus even under the masking conditions, and is simply responding to that.

Can we be sure that the A-figure was in fact subliminal, in the sense that the A-figure was not even partly discriminable? The present experiment did not explicitly aim to explore this possibility, and we must grant that our findings do not settle the matter. None of the *S*s gave any indication that they were aware of another form other than the B-figure, but of course our method of report did not completely rule out this possibility.[8] In another study, we propose to obtain discrimination thresholds under a variety of conditions designed to produce optimal circumstances for at least the partial recognition of a stimulus. Thresholds obtained in this way will then be used to set stimulus exposures at a level where we can safely presume that not even partial discrimination is possible. We shall then explore the behavioral effects of such subliminal exposures in a manner patterned after our present procedure as well as that of Fisher (8, 9) in his closely related work.

SUMMARY

A method is presented for studying cognitive effects of stimuli that do not give rise to conscious percepts. The *S* is shown a drawing of a person ambiguous as to sex, in a mirror tachistoscope, for .33 second. This exposure is preceded by a briefly exposed stimulus (approximately .027 second) which is not consciously perceived. The first, subliminal presentation is called *A-figure*, the second one *B- figure*. In this experiment, the B-figure was always a sexually ambiguous drawing of a person.

In a balanced design using 24 male medical students, this technique was used to study the effects of four A-figures—a male symbol, a female symbol, male genitals, and female genitals—on three different B-figures. In the first session, *S* was shown a B-figure preceded by a subliminal A-figure; he then drew the B-figure and rated it on 69 four-point adjective scales. Similar ratings and drawings were obtained after each exposure of a B-figure. In the second session, an A-figure opposite in sex was paired subliminally with the same B-figure. Next, another B-figure was shown preceded by supraliminal exposure (.33 second) of another A-figure. Finally, ascending thresholds for the previously subliminal A-figures alone were also obtained, and *S*s gave associations to the symbolic A-figures. Control exposures without any preceding A-figures were also rated and drawn by all *S*s.

Effects of the subliminal figures were studied by comparing each *S*'s check-list rating of the B-figure with a group consensus of the appropriate A-figure. Drawings were rated for masculinity. Effects of the subliminal *genitals* were highly consistent on both check list and drawing, with *S*s tending either to incorporate or to exclude attributes of the subliminal A-figure in their impressions of the B-figure. Subliminal *symbols* were significantly effective in influencing *S*'s drawing of the B-figure; but this effect was not correlated with response on the check list. Recognition thresholds for the genital A-figures were related to check-list responsiveness, while thresholds for the symbolic A-figures were related only to responsiveness on the drawings.

REFERENCES

1. ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, **54**, 383–405.
2. BINET, A. *Alterations of personality.* (Trans. by Helen G. Baldwin) New York: Appleton, 1896.
3. BLUM, G. S. Perceptual defense revisited. *J. abnorm. soc. Psychol.*, 1955, **51**, 24–29.
4. BRICKER, P. D., & CHAPANIS, A. Do incorrectly perceived tachistoscopic stimuli convey some information? *Psychol. Rev.*, 1953, **60**, 181–188.
5. CHEATHAM, P. G. Visual perceptual latency as a function of stimulus brightness and contour shape. *J. exp. Psychol.*, 1952, **43**, 369–380.

---

[8] The fact that one *S*, whose data were subsequently discarded, did report awareness of "something else" preceding the picture, shows that *S*s could report such experiences if they had them.

6. ERIKSEN, C. W. Perceptual defense as a function of unacceptable need. *J. abnorm. soc. Psychol.*, 1951, **46**, 557–565.
7. ERIKSEN, C. W. Subception: Fact or artifact? *Psychol. Rev.*, 1956, **63**, 74–80.
8. FISHER, C. Dreams and perception. *J. Amer. psychoanal. Ass.*, 1954, **2**, 389–445.
9. FISHER, C. Dreams, images and perception: A study of unconscious-preconscious relationships. *J. Amer. psychoanal. Ass.*, 1956, **4**, 5–47.
10. FREUD, S. *The interpretation of dreams.* Vols. IV & V. (Trans. by J. Strachey) London: Hogarth Press, 1953.
11. HELSON, H. Perception and personality—a critique of recent experimental literature. *USAF Sch. Aviat. Med. Proj. Rep.* 1953, Project #21-0202-0007 (Report #1).
12. KARLIN, L. The New York University tachistoscope. *Amer. J. Psychol.*, 1955, **68**, 462–466.
13. KRIS, E. On preconscious mental processes. *Psychoanal. Quart.*, 1950, **19**, 510–560.
14. LACEY, J. I., & SMITH, R. L. Conditioning and generalization of unconscious anxiety. *Science*, 1954, **120**, 1045–1052.
15. LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: A study of subception. *Psychol. Rev.*, 1951, **48**, 113–122.
16. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
17. MILLER, J. G. Unconscious processes and perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: An approach to personality.* New York: Ronald, 1951.
18. OSGOOD, C. *The measurement of meaning.* Champaign: Univer. of Illinois Press, 1958.
19. POETZL, O. Experimentelle erregte Traumbilder in ihren Beziehungen zum indirekten Sehen. *Zfl. ges. Neurol. Psychiat.*, 1917, **37**, 278–349.
20. RAPAPORT, D. *Emotions and memory.* Baltimore: Williams & Wilkins, 1942.
21. SCHILDER, P. *Medical psychology.* (Trans. by D. Rapaport) New York: International Universities Press, 1953.
22. SHEVRIN, H., & LUBORSKY, L. The measurement of preconscious perception in dreams and images: An investigation of the Poetzl phenomenon. *J. abnorm. soc. Psychol.*, 1958, **56**, 285–294.
23. VOOR, J. H. Subliminal perception and subception. *J. Psychol.*, 1956, **41**, 437–458.
24. WALKER, H. M., & LEV, J. *Statistical inference.* New York: Henry Holt, 1953.
25. WERNER, H. Studies on contour: I. Qualitative analyses. *Amer. J. Psychol.*, 1935, **47**, 40–64.

Received July 26, 1957.

A STUDY OF ANXIETY AND LEARNING IN CHILDREN[1]

RICHARD R. WAITE, SEYMOUR B. SARASON, FREDERICK F. LIGHTHALL, AND KENNETH S. DAVIDSON

Yale University

THIS study focuses on differences between high anxious (HA) and low anxious (LA) children in an experimental learning situation. Two recent reports described test anxiety and general anxiety scales for children[2] and presented evidence relevant to the problem of validity (1, 6). The present study is similarly concerned with the validity of the children's scales. Do groups defined in terms of scores on the scales perform differently in an experimental learning situation? It was expected that any such over-all differences would be in the same direction as that found in college studies of test anxiety, LA subjects doing better than HA subjects (1, 2, 5). Another objective was the investigation of the effects of differential instructions (success, failure, and neutral) on HA and LA children. Again our expectations were derived from the results of the college studies in which LA subjects did best when given either success or failure instructions, whereas HA groups performed optimally under neutral conditions.

The conviction with which we hypothesized the above differences was tempered by a number of facts: (*a*) our present subjects were obviously much younger than college students, (*b*) our college students were males, whereas in the present study both sexes were used, and (*c*) the intelligence range of the college students was restricted, whereas in a grade school such limitations are much less. In the light of these and other considerations, it was felt that findings similar to those obtained with college students would not only be relevant to the validity of the scales but also to an understanding of the developmental aspects of the anxious reaction.

METHOD

Subjects

The children in this study came from a group of 747 pupils in Grades 2 through 5 to whom the test anxiety and general anxiety scales had been administered in the previous grade. All pupils had taken the Otis Alpha Intelligence Scale as well. Excluded were those regarded by their teachers as arithmetic, reading, or behavior problems, and those whose parents were separated or divorced. The *S*s whose test and general anxiety scores fell in the upper quartile of the distributions were considered HA, and those whose scores fell in the lower quartile were termed LA. Twenty-four pairs of children were then selected by matching LA *S*s with HA *S*s in terms of grade, sex, and Otis Alpha scores. Three such pairs were obtained for each grade-sex group, and one of these pairs was assigned to each of the three instruction groups—neutral, success, and failure. Thus, the design was a 2 (anxiety) by 3 (instructions) by 2 (sex) by 4 (grade) arrangement, with one *S* in a cell. The matching procedure, even with such a large group of *S*s from which to draw pairs, was found to be difficult. It was evident that to have had more than one person in a cell, while at the same time including all the specified variables, would have required many more than the original 747 children.

Procedure

The *E*, who did not know which *S*s were classified as HA or LA, worked with them individually in the public schools. After bringing *S* from the classroom to the testing room, an introductory conversation took place during which the establishment of good rapport was attempted. Then *E* asked the *S* to tell a story to a TAT card,[3] after which the following instructions were given.

> I have a test I'd like you to take today. This is a test that I have given to lots of children in all of the schools in Milford. I've given the test to children in all the grades, so that I know just how well a first grader can do, how well a second grader can do, a third grader, a fourth, fifth, sixth, and seventh grader can do. Of course, a second grader does better than a first grader, a fourth grader does better than a third grader, and so on, because he's older, right? Now, your teacher and I want to find out how well *you* can do on the test—whether you do as well as other [4th] graders, or only as well as [3rd or 2nd] graders, or as good as [5th or 6th] graders.
>
> Now, I have 12 cards here with a figure and a number on each. There's a different figure and a different number on each card. I'm going to show you the cards one at a time, and I want you to try to remember the number that goes with each figure. When I've shown you all the cards once, I'll give you a piece of paper with the figures on it and you try to write the correct number under each of the figures. Then I'll show you the cards again and ask you to

---

[1] The project of which this study is a part is being supported by a grant from the United States Public Health Service, (M-712).

[2] Copies of these two scales are available for research purposes only.

[3] We wish to express our thanks to Robert P. Abelson for his most helpful advice on the statistical analyses.

write the numbers again. We'll do this six times in all, so you'll have plenty of chances. O.K.?

The cards were presented at two-second intervals, and *S*s were allowed as much time as they desired in writing the numbers. While they were filling out the "answer sheets" for each trial, *E* pretended that he was correcting the results of the previous trial. Upon completion of the sixth trial, *E* "added up the results" and then made the following statements, appropriate to the instruction group of which the *S* was a member.

> *Success S*s. "You did as well as kids in the [2 grades above *S*'s actual grade] grade. You can be very proud. Now I'm going to give you another test just like this one. This one has the same numbers, but different figures."
>
> *Failure S*s. "You didn't do well at all on this test. You only did as well as [2 grades below *S*'s actual grade] graders. You should have done much better. Now I'm going to give you another test just like this one. This one has the same numbers, but different figures."
>
> *Neutral S*s. "Now I'm going to give you another test just like this one. This one has the same numbers, but different figures."

The second set of six trials was then given, using a new set of stimulus cards. Then all *S*s were assured that they had done well, with extra attention being given those children who had received failure instructions, in order to leave them at ease. After *S* had returned to his classroom, *E* filled out a behavioral check list which included items such as nail-biting, scratching, apparent motivation, reaction to instructions, etc., and wrote a brief report on his general observations during the testing situation.

RESULTS

Figure 1 presents the over-all results of the two anxiety groups. For each group, mean scores on each trial were calculated and plotted to provide group learning curves. At the beginning of the first task, the group means are identical, but by the conclusion of this task, the LA group is decidedly superior in performance. When the same data are divided into sex-anxiety groups, the differences in anxiety-group means hold up better within males than within females. In general, the girls appear to do better than the boys.

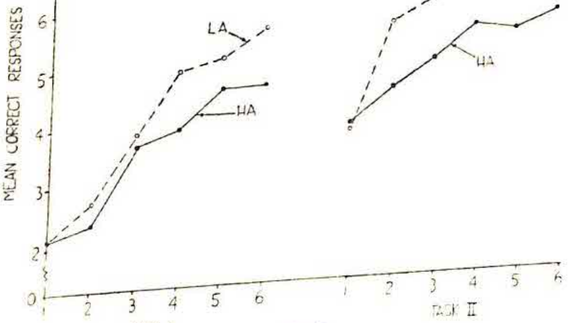

FIG. 1. LEARNING CURVES FOR THE TWO ANXIETY GROUPS BEFORE AND AFTER INSTRUCTIONS

Individual scores were treated in the following manner: First, for each *S* a mean score over each set of six trials was computed. Second, the score of the first trial in each of the two sets of trials was multiplied by −5, the second by −3, the third by −1, the fourth by 1, the fifth by 3, and the sixth by 5. These products were then summed to give a comparison number which can be considered an index of the linear component of the slope of the individual's learning curve, or an estimate of his rate of learning. For each *S*, then, there are two mean scores (one for each set of six trials) and two slope scores.

Since there were no differential instructions given on Task I, it was possible to carry out an analysis of variance on the first set of trials, collapsing on instructions to leave three *S*s in a cell. Two analyses were done, one for slopes and one for means. In terms of means, there is no significant difference between the LA and HA groups. The grade effect is significant, but the data show that instead of an increase in mean scores from grade to grade as we would expect, there is a drop from second to third grade, and then an increase to fourth and to fifth grade. The analysis of slopes shows parallel results. See Footnote 3.

Next, analyses of the Task II data were carried out. In this case it was necessary to construct an error term from the grade-by-

TABLE 1
ANALYSIS OF VARIANCE OF TASK II MEAN SCORES

Source	*df*	Mean Square	*F*	*p*
Sex	1	18.54	4.14	.10
Instructions	2	.40	—	—
Grade	3	50.85	11.43	.01
Anxiety	1	18.13	4.08	.10
S × I	2	4.06	—	—
S × G	3	10.42	2.34	.10
S × A	1	1.63	—	—
I × A	2	4.25	—	—
G × A	3	14.76	3.32	.05
S × I × A	2	1.15	—	—
S × G × A	3	3.55	—	—
Error	24	4.45		

instructions interactions; grade was used because it was thought to be a relatively unimportant variable in terms of the purposes of this study. The analysis of means (Table 1) again shows significant grade effects. Anxiety and sex are significant at the .10 level, while the grade-by-anxiety interaction is significant at the .05 level. The data show that the girls did somewhat better than the boys, that the grade effect is essentially linear, and that LA *S*s did better than HA *S*s. Furthermore, the significance of the grade-by-anxiety interaction seems due, to a great extent, to the first-grade HA girls scoring much higher than one would expect (higher than fourth-grade HA or LA girls). The analysis of slopes (Table 2), on the other hand, gives much clearer results. Only anxiety is significant and here at the .01 level. LA *S*s learned faster than HA *S*s. Of note is the nonsignificance of the instructions effect in both analyses.

Since the differences in slopes of the anxiety groups may have been a carry-over from the first task, sufficiently increased through practice to reach significance, an analysis of covariance was performed on the Task II slopes, controlling for Task I slopes. Table 3 shows that the difference in Task II slopes remains significant at the .05 level.

The check lists which were obtained at the end of each experimental session were totalled for each anxiety group. No differences in any of the items were found.

TABLE 2
ANALYSIS OF VARIANCE OF TASK II SLOPE SCORES

Source	*df*	Mean Square	*F*	*p*
Sex	1	295.02	—	—
Instructions	2	207.06	—	—
Grade	3	569.35	1.54	—
Anxiety	1	2041.02	5.52	.01
S × I	2	116.90	—	—
S × G	3	246.58	—	—
S × A	1	792.19	2.14	—
I × A	2	585.27	—	—
G × A	3	299.58	—	—
S × I × A	2	324.19	—	—
S × G × A	3	341.19	—	—
Error	24	369.22		

TABLE 3
ANALYSIS OF COVARIANCE OF TASK II SLOPE SCORES

Source	*df*	Adjusted Mean Square	*F*	*p*
Anxiety	1	1255.98	11.33	.05
Within	45	110.87		
Total	46			

DISCUSSION

The most interesting results as far as the major purpose of the experiment is concerned are the significant differences between the anxiety groups. Since the groups were at least roughly matched on other factors which operate in learning situations with children—sex, age, and intelligence—we can only assume that whatever it is that the anxiety scales measure is highly related to the differences in performance. As expected, the direction of the results coincides with that obtained in similar studies with college students. Thus, we can conclude that the experiment has further demonstrated the validity of the children's scales.

The lack of any significant difference between the anxiety groups on the first task is an aspect of the results about which we can only conjecture. One explanation may lie in the nature of the two learning tasks. While they were alike in manner of presentation, size, other formal characteristics of the stimuli, and the nature of the learning involved, there was a slightly higher intralist similarity on the first task. Consequently, it would be expected that the learning involved on the first task would be more difficult, both because of the higher intralist similarity and the initial position of the task. The net result may be thought of in terms of a general depression of scores, cutting down the possibility of differences.

An alternative explanation involves viewing the effect of instructions in a different way. In this study it apparently made no difference what was said to the *S*s. Perhaps it was the case that any statement about another test, no matter what the specific verbalization, cued off different reactions in the two anxiety groups, reactions which ultimately affected performance. The present study provides no information relevant to the veracity of this hypothesis, however.

One other problem deserves mention and brief discussion. One of our interests in anxiety, particularly test anxiety, concerns both its interfering and facilitating characteristics in

Bureau Ednl. Research
DAVID HA ING COLLEGE
Dated

different kinds of learning situations. In the present study, the differences between HA and LA *S*s suggest the presence of interfering factors in the former group. Whatever the degree of interference, its significance derives in large part from the fact that the two groups were matched in terms of an intelligence test score, the assumption underlying such matching being that the two groups are approximately equal in general intelligence and should therefore perform similarly in the learning situation. However, since our studies of children and college students (4) have consistently a significant *negative* relationship between IQ score and test anxiety,[4] it should follow that matching HA and LA *S*s on IQ, as was done here, involves matching on a measure which is already influenced by anxiety. It could be argued, therefore, that when HA and LA groups are so matched, the "true" score of the former is likely to be higher than that of the latter. From this standpoint, one may speculate that the performance of the HA *S*s in the present study reflects a degree of interference that is greater than a comparison of the learning curves of HA and LA *S*s indicates; i.e., the degree of discrepancy between potential and performance is greater among HA *S*s than the learning curves suggest.

SUMMARY

A general anxiety scale and a test anxiety scale were administered to 747 grade school children. Out of this group, 24 HA–LA pairs of subjects were given two modified paired-associate learning tasks, separated by neutral, failure, and success instructions. The results showed no differences due to instructions, but LA subjects performed better than HA subjects on the second task. Alternative explanations for the lack of differences on the first task were offered. The study was interpreted as positive evidence for the validity of the anxiety scales.

REFERENCES

1. GORDON, E. A. Self esteem, need for achievement, test anxiety and performance. Unpublished doctoral dissertation, Yale Univer., 1955.
2. MANDLER, G., & SARASON, S. B. A study of anxiety and learning *J. abnorm. soc. Psychol.*, 1952, **47,** 166–173.
3. SARASON, S. B., DAVIDSON, K. S., LIGHTHALL, F. F., & WAITE R. R. A test anxiety scale for children. *Child Develpm.*, in press.
4. SARASON, S. B., & MANDLER, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47,** 810–817.
5. SARASON, S. B., MANDLER, G., & CRAIGHILL, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47,** 561–565.
6. SARNOFF, I., LIGHTHALL, F. F., WAITE, R. R., DAVIDSON, K. S., & SARASON, S. B. A cross-cultural study of anxiety amongst American and English school children. *J. educ. Psychol.*, 1958, **49,** 129–136.

Received July 29, 1957.

---

[4] We have computed correlations between IQ and test anxiety for several hundred classrooms. Approximately 90% of these correlations were negative.

THE EFFECTS OF DIFFERENT SOCIAL FEEDBACK CONDITIONS UPON PERFORMANCE IN DYADIC TEAMS[1]

SEYMOUR ROSENBERG[2] AND ROBERT L. HALL[2]
USAF Personnel and Training Research Center

IN THE present study, one of three types of feedback stimuli was presented to a group member following a response. They are: (*a*) information based upon the error magnitude of the subject's own response only, termed "direct" values. This type of feedback stimulus is typically used in individual learning studies to differentiate a particular range of response magnitudes; (*b*) information that combines in a simple additive function the error magnitudes of both the subject's own response and that of his teammate, termed "confounded" values; (*c*) information based upon the error magnitude of a teammate's response only, termed "other's" values. These three feedback conditions may be viewed as comprising two extremes and an intermediate point of a continuum, which specifies the fractional weight given to a subject's response relative to the fractional weight given to his teammate's response in the feedback stimulus to the *subject*. The feedback value to the subject is "direct" when his response weight in the feedback is 1.0 and his teammate's response weight is 0.0 (one extreme of the continuum). The subject's feedback value is "other's" when his own response weight in his feedback is 0.0 and his teammate's is 1.00 (other extreme of the continuum). The subject's feedback value is "confounded" when his feedback values contain nonzero weighing of both his own and his teammate's response. The present paper will describe the effects of these feedback conditions upon individual and team measures, and upon changes in these measures over a series of discrete trials.

The stimulus event following a response has been variously termed: feedback, knowledge of performance, knowledge of results, reinforcement, terminating stimulus, etc. (e.g., 2, 3, 9, 11). There is little disagreement about the importance of such general parameters of this stimulus event as its frequency and periodicity, coarseness of measurement, delay, manner of presentation, type and amount of informational error, number of trials, etc. However, these parameters have been studied almost exclusively in experiments using individual subjects. Since there has been very little systematic attempt either to determine the values of such parameters particularly relevant to interpersonal settings, or to formulate variables unique to these settings, the research to date is incomplete or, at best, difficult to apply to interpersonal settings. The application or extension of parameters to interpersonal contexts presents a number of interesting and novel problems. For example, when a stimulus event following a response by Person A is some joint function of his behavior and that of a Person B, it is difficult to specify in quantified terms the stimulation of A from knowledge obtained in individual research and, further, to specify changes in stimulus parameters that occur over time. In attempting to apply current knowledge to interpersonal settings, the investigator may also encounter the need to formulate novel variables. For example, one class of variables unique to interpersonal settings is variations in the mathematical function that expresses the way the responses of Persons A and B are combined to produce a stimulus event to one or both persons. Distributional variables of any sort, i.e., combinations of treatments of group members, are also novel. Neither in social psychology nor in learning psychology has there been any systematic experimental analysis of social and learning variables simultaneously. This study is one of the early experiments in such a program. This program was described in some detail in a paper by Hall and Rosenberg.[3]

---

[1] This report is based on work done under ARDC Project No. 7713, Task No. 27104, in support of the research and development program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

[2] Seymour Rosenberg is now at the University of Kansas. Robert L. Hall is now at the University of Minnesota.

[3] Hall, R. L., & Rosenberg, S. An S-R analysis of social interaction. Paper presented at American Psychological Association meetings in New York, September 1957.

METHOD

Subjects. The subjects (*S*s) were 54 male Air Force basic trainees whose modal age was about 17 years. They were assigned as a unit for use as *S*s in psychological research. *S*s from such units were excluded from the experiment if their Armed Forces Qualification Test (AFQT) score was below the 65th percentile rank, which is roughly equivalent to a Stanford Binet IQ of 108 (4). *S*s were used in pairs, randomly assembled.

Apparatus. The two *S*s were seated in identical adjacent booths. The walls of the booth were sufficiently high so that *S*s could not see each other. One wall of each booth consisted of the panel and table shown in Figure 1. Each *S* faced a separate panel. A micrometer was mounted behind the shield so that the knob at the beginning of each trial projected ¼ inch from the face of the panel. The *S* could touch only the knob portion of the micrometer. He could turn the knob by reaching under the shield, but he received no visual cues nor special tactual cues from the knob to indicate the amount it was turned. The knob could only be turned to the left. The experimenter (*E*), seated on the other side of the panel, could read the micrometer scale, which is a linear measure of the amount the knob is turned. Also, on the panel in front of each *S*, there was a linear scale extending from zero in the center to 80 in each direction, indicating "size of error." A movable pointer above the scale displayed a score to the *S* after each trial. The scale was concealed by a hinged shield during the *S*s' response period and the feedback delay period. The scale also contained "off the scale" marks at each end. At the center of the scale an interval of 10 units was marked off by two red lines and was labeled "try to get pointer within red lines for good score." No cues were present on the scale to indicate whether error scores to the left of zero on the scale indicated too large or too small a response magnitude. For counterbalancing, excessive response magnitudes were shown to the left of zero for half of the *S*s and to the right of zero for the other half of the *S*s.

Procedure. Each pair of *S*s was given the same instructions, followed by 50 response-feedback trials. Each member of a pair was independently assigned to one of the three feedback conditions for all 50 trials. To satisfy a convenient analysis of variance design to be described later, certain restrictions were imposed upon the numbers of *S*s assigned to each condition.

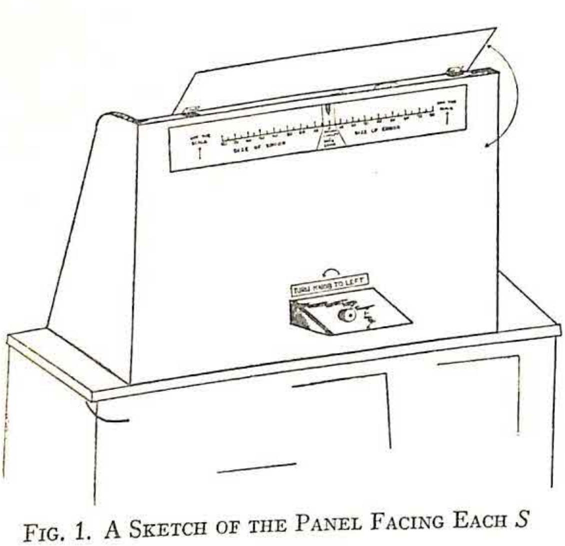

FIG. 1. A SKETCH OF THE PANEL FACING EACH *S*

After the two *S*s were seated in adjacent booths, a few minutes were spent by the *E* "setting them at ease." The identical panels, one for each *S*, were then put in place. The following tape-recorded instructions were played:

> The Air Force needs to know more about how well people can do certain tasks. The experiment in which you are to take part is a study of just one such task. It is a simple task of turning a knob the right amount. Listen carefully to these instructions and the experimenter will point out things as they are mentioned. For this experiment a panel will be placed beside each of you. On the front of the panel you see a shield and underneath the shield, where you cannot see it, is a small knob. Do not look under the shield. You are expected to work without seeing the knob. Your job is to learn to turn your knob just the right amount. You will have a number of chances to do this. To help you, you will be shown, after each trial, the error in the score that you made. Your error score will be shown you on the scoreboard at the top of your panel. The movable pointer, which is moving now, tells you how much you missed. To make a *good* score, try each time to get the pointer within the two red lines. Always turn your knob to the left as shown by the small sign tacked on your panel. A buzzer will be used to tell you when to start and stop. The buzzer sounds like this. [Buzzer is sounded.] When the buzzer starts, you start to make your setting. As soon as you have made the setting you think is correct, take your hand away from the knob. When the buzzer stops, you must take your hand away whether you are finished making your setting or not. You will find that you have plenty of time without rushing. A few seconds after the buzzer stops you will be shown on your scoreboard the error in the setting you just made. Shortly after that, you will hear the buzzer again. This means that you are to start again and try once more to make a good score. After each trial your knob is set back by us. Each time the buzzer starts, you start over and try again to make a good score. Now we are ready to start. Do not touch your knob until you hear the buzzer.

It should be noted that the *S*s were given no cues that they might be working as a team. They did not talk to each other. To control for inadvertent cues by *E* to *S*s in different treatments, each pair of *S*s was assigned to one of the treatments only *after S*s were seated in the closed booths and instructed.

The procedure for one complete trial consisted of three automatically timed phases: (*a*) a response period of eight seconds, indicated to the *S*s by a mild buzzer, during which the knob could be turned; (*b*) a delay of eight seconds during which *E* recorded response magnitude and moved the scale pointer (still concealed from each *S*) to its appropriate value for that trial; (*c*) exposure of the pointer position to the *S* by lifting the hinged shield. During this interval the micrometer was reset by *E* to its original position. The shield was

TABLE 1

SUMMARY OF THE STATISTICAL ANALYSES OF THE THREE RESPONSE MEASURES

Source	*df* of each Analysis			Individual Accuracy $A_i = \log (\lvert X - 100 \rvert + 1)$		Team Accuracy $A_t = \log \lvert X + Y - 200 \rvert + 1)$		Differentiation $D_t = \log (X - Y) + 1$		Error Term for *F*
				Mean Sq.	*F*	Mean Sq.	*F*	Mean Sq.	*F*	
1. Teams	26			83009		70602		59782		
A. Rows (Feedback cond.)		2		620276	17.32***	174340[a]	5.46**	209026[a]	3.46*	1D
B. Columns (Feedback cond.)		2		10784	.30	284080		121978		
C. Rows × Columns		4		63434	1.77	40915	.98	7984	.17	1D
D. Within Cells		18		35817		41952		47799		
2. Trials	49			2312		4102		3041		
3. Teams × Trials	1274			1405		1559		1559		
A. Rows × Trials		98		2089	1.54**	1242	1.09[a]	1439	.89[a]	3D
a. Linear			2	22418	4.51*	3854	2.38	16513	.92	3Da
b. Quadratic			2	10372	4.19*	4438	1.56	2452	1.09	3Db
c. Residual			94	1480	1.18	1119	.90	1097	.87	3Dc
B. Columns × Trials		98		1025	.75	2120	—[a]	1515	—[a]	3D
a. Linear			2	1274	.26	35148	—[a]	5046	—[a]	3Da
b. Quadratic			2	500	.20	5582	—[a]	6890	—[a]	3Db
c. Residual			94	1031	.82	1343	—[a]	1325	—[a]	3Dc
C. Rows × Columns × Trials		196		1463	1.08	1486	.96	1448	.87	3D
a. Linear			2	6751	1.36	7806	.95	9863	.84	3Da
b. Quadratic			2	10263	4.15*	2706	.84	6140	1.43	3Db
c. Residual			94	1163	.92	1326	.97	1169	.84	3Dc
D. Within Cells × Trials		882		1358		1548		1660		
a. Linear			18	4966		8201		11753		
b. Quadratic			18	2473		3212		4280		
c. Residual			846	1258		1371		1389		

[a] For each of the two team measures, A_t and D_t, rows and columns mean squares refer to exactly the same effects. Therefore, these two sources of variance were pooled within each analysis of variance and a single *F* ratio computed. The same procedure was adapted for the "rows × trials" and "column × trials" variances. Note that the numerator of such *F*s has associated with it four *df*.

$*P < .05$; $**P < .01$; $***P < .001$.

lowered after 5–10 seconds and the response interval for a new trial was started.

Response magnitude for each *S* during each trial was read by the *E* directly from the micrometer scale. Twenty-five scale units equals one complete knob rotation. Where an *S* was receiving "direct" feedback (his response magnitude only), a perfect score (goal score) was 100 micrometer units (four knob turns). Where an *S* was receiving "confounded" feedback (the combination of his own and his teammate's response magnitude), a score to the *S* was computed by averaging the response magnitudes of the two persons. An average of 100 units was a perfect score. An *S* receiving "other's" feedback (his teammate's response magnitude) received a perfect score when his teammate made four knob turns or 100 micrometer units.

Measurement and statistical design. Three measures were computed for each trial and transformed to logarithms to reduce skewness:[4] (*a*) "individual accuracy," $A_i = \log(\lvert X - 100 \rvert + 1)$, where X is response magnitude, in micrometer units of one *S*; (*b*) "team accuracy," $A_t = \log (\lvert X + Y - 200 \rvert + 1)$, where X and Y are response magnitudes of the two *S*s in micrometer units; (*c*) "response (role) differentiation," $D_t = \log (\lvert X - Y \rvert + 1)$.

Each of the three measures was subjected to a similar analysis of variance. The design consists of three rows corresponding to the three feedback conditions to which one *S* could be assigned and three columns corresponding to the same three feedback conditions to which his teammate could be assigned. Thus, there are nine row-column cells. Three pairs of *S*s were assigned to each of these nine cells (combinations of feedback conditions). Fifty trials constitutes a third criterion of classification in the analysis.

For the A_i measure, one member of each dyad was randomly selected. Data from the second member were not analyzed. Thus, the same team is never represented twice in the analysis of A_i. Two analyses of variance of A_i were also avoided because of the complex experimental dependency between the two *S*s in most of the combinations of feedback conditions. The rows represent the feedback condition of the member whose A_i is contained in the cell. The columns represent the condition of his partner.

When either team measure (A_t or D_t) is used in this statistical design, row and column classifications bear the same relationship to the dependent variable, i.e., represent the feedback condition for one of the team members. Put in another way, the row variable is exactly the same as the column variable. In such a design, combinations of experimental conditions on one side of one of the diagonals are also found on the other side of that diagonal. Data must be arbitrarily distributed between cell-pairs whose conditions are the same.

Two orthogonal polynomial components were extracted from interaction effects containing trials. The components were the "linear" and "quadratic" and were used to test the presence of differential rates of development of a measure with trials. These components supplement the test for over-all trials effects since the latter is too general a test of trend and fails to distinguish simple trend types. The rationale and method of extraction of a linear component in a trend analysis is described in several places in the psychological literature (1, 8), and the more generalized procedure for extraction of orthogonal polynomials is described by Grant (5).

---

[4] The value, 1, was added to the absolute value in each of the three measures to avoid the log of 0, which has an indeterminate value.

RESULTS

Analyses of variance of the three measures are summarized in Table 1. Note that row and column effects are pooled in the analysis of A_t and D_t since the effects tested are the same. For A_i, the effects refer to two different operations with respect to the dependent variable, i.e., an *S*'s response measure classified by his own feedback conditions (rows) and that of his partner (columns).

In the case of A_i, different feedback conditions of the *S*s result in significant differences in average level of his accuracy over the 50 trials (row effects), in over-all trend differences (interaction of rows with trials), and in linear and quadratic trend components. Figure 2(A) illustrates these differences. Note that on the ordinate, the smaller A_i values are associated with greater accuracy, and that on the abscissa the trials are grouped in sets of five. The graph points up a rather simple positive relationship between individual accuracy and contribution of the *S* to his own feedback. In striking contrast, different feedback conditions of *S*'s partner (column effects) do not result in any significant differences in *S*s' accuracy. This is apparent from Figure 2(B) which illustrates no consistent divergencies among groups. Figures 2(A) and 2(B) would not illustrate the presence of any unique effects produced by particular combinations of an *S*'s condition and that of his partner, if any existed. For example, if the combination of "other's–other's" conditions resulted in an exaggerated degree of inaccuracy for *S* (over and above that accounted for by the main effects), these graphs would fail to show this effect. However, such an effect would be reflected in "rows × columns" and/or "rows × columns × trials" interactions in the A_i analysis. The results in Table 1 fail to support, with any statistical significance, the presence of such combinatorial effects except in the isolated instances of a quadratic component in Source 3C.

An examination of both team measures, A_t and D_t, reveals a rather consistent picture. For both measures, differences in the feedback conditions yield significant differences on the average of the 50 trials (Sources: Sum of 1A and 1B). Differences in trend, however, are not revealed (Sources: Sum of 3A and 3B) unless one is willing to use a more liberal confidence level (or another error term) as a basis for rejecting null hypothesis. In the case of A_t, the *F* value of 2.38 (*df*, 4, 18) for linear differences in trend falls between the .05 and .10 levels. Figures 3(A) and 3(B) illustrate the A_t data for rows and columns, respectively, and do suggest the existence of trend differences among conditions. The "replication" between row effects and column effects does not appear to be perfect, but no reversals are apparent. It is interesting to note, in Figure 3(B), that confounded feedback results in somewhat bet-

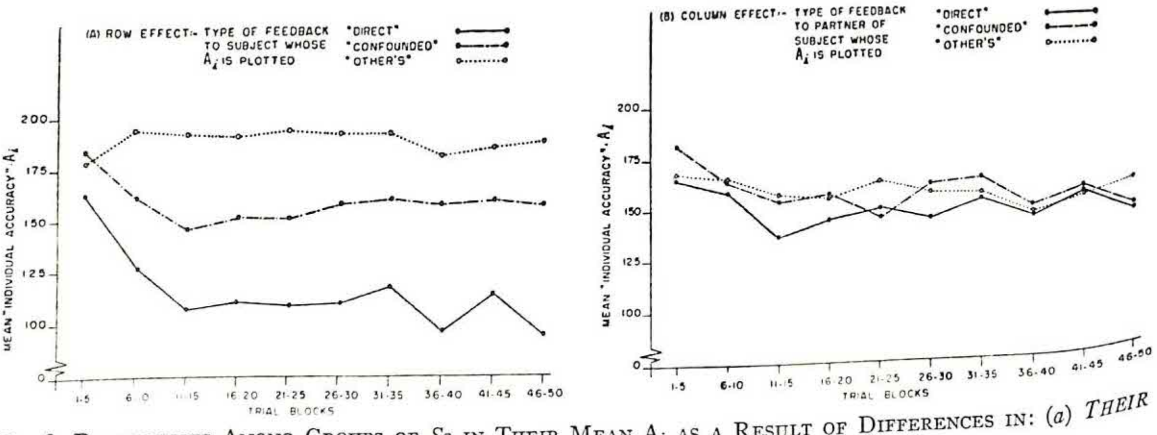

FIG. 2. DIFFERENCES AMONG GROUPS OF *S*S IN THEIR MEAN A_i AS A RESULT OF DIFFERENCES IN: (*a*) *THEIR OWN* FEEDBACK CONDITIONS; (*b*) FEEDBACK CONDITIONS OF THEIR *PARTNERS*

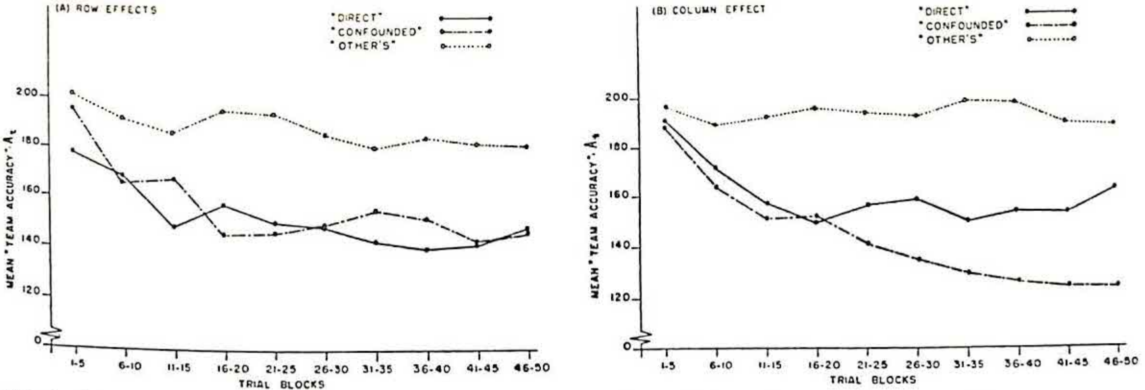

FIG. 3. DIFFERENCES AMONG GROUPS OF SUBJECT-PAIRS IN THEIR MEAN A_t AS A RESULT OF DIFFERENCES IN FEEDBACK CONDITIONS TO: (*a*) ONE MEMBER OF THE PAIR; (*b*) THE SECOND MEMBER OF THE PAIR

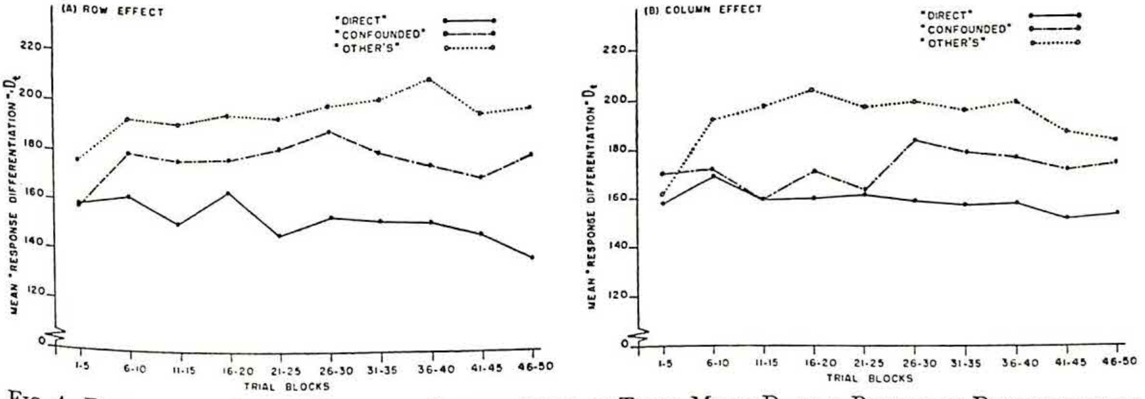

FIG. 4. DIFFERENCES AMONG GROUPS OF SUBJECT-PAIRS IN THEIR MEAN D_t AS A RESULT OF DIFFERENCES IN FEEDBACK CONDITIONS TO: (*a*) ONE MEMBER OF THE PAIR; (*b*) THE SECOND MEMBER OF THE PAIR

ter team accuracy than direct feedback to each *S*. Figures 4(A) and 4(B) present for rows and columns, respectively, the D_t measure, the ordinate indicating the mean magnitude of difference between teammates' responses. "Replication" is again somewhat imperfect although no serious contradictions are apparent. Teams containing an *S* who received as feedback only his teammate's response showed greatest D_t, while teams with an *S* receiving "direct" feedback showed the least D_t.

Finally, it is again important to note that Figures 3 and 4 would not illustrate unique effects in A_t and D_t produced by particular combinations of feedback conditions to which pairs of *S*s were exposed. However, the analyses of the two team measures in Table 1 fail to support the statistical significance of any such combinational effects since none of the "rows × columns" or "rows × columns × trials" interactions are significant. Team accuracy and response differentiation values (as quantified in this study) are apparently determinable from an additive weighting of each of the two feedback conditions of the *S*s.

DISCUSSION

As noted previously, the three feedback conditions in the study can be viewed as points on a continuum. The continuum is describable as the proportion of contribution of an *S*'s own response to his feedback, relative to the contribution of his teammate's response. The continuum may be expressed as the ratio of the weights accorded responses of the two *S*s in a simple weighted average; that is, the size of the ratio, k_1/k_2, in the simple function:

$$\text{Feedback} = k_1X + k_2Y - G$$

where X and Y are response magnitudes of two *S*s, G is goal score, and $k_1 + k_2 = 1$. The three feedback conditions would then be expressed:

a. Feedback $= (1)(X) + (0)(Y) - 100 = X - 100$, for Subject X, or Feedback = $(0)(X) + (1)(Y) - 100 = Y - 100$, for Subject Y, in "direct" feedback.

b. Feedback = $(.5)(X) + (.5)(Y) - 100 = .5X + .5Y - 100$ for both *S*s, in "confounded" feedback.

c. Feedback = $(0)(X) + (1)(Y) - 100 = Y - 100$, for Subject X, or Feedback = $(1)(X) + (0)(Y) - 100 = X - 100$, for Subject Y, in "other's" feedback.

The findings concerning A_i (individual accuracy) are rather simply summarized in terms of its relation to the k_1/k_2 ratio. Both average magnitude and rate of development of A_i are related monotonically to the magnitude of k_1/k_2 where k_1 and k_2 are the constants applied to the feedback function of the *S*. No relationship was found between A_i of an *S* and the k_1/k_2 ratio of his teammate's feedback function.

The use of this ratio to summarize succinctly the relationships discovered early in this research program imply little about its future utility. A more articulated theory may not use this variable. One could also describe, for example, the feedback conditions along a continuum or continua which express the "quality of the feedback." Quality of feedback may refer to the size *and* consistency of the covariance between feedback values and the response magnitudes of an *S* (i.e., reliability of feedback as a measure of response magnitude). The correlation coefficient may be used as a measure of this covariance, although more comprehensive measurement techniques are probably required.[5] Quality of feedback may also refer to any constant difference in magnitude between a feedback value and a response value. In other words, quality is reduced by both variable and constant errors in the feedback. It may be noted that "quality of feedback" is oriented toward individual feedback. Such an orientation points up, however, a number of interesting feedback variables that are likely to arise in interpersonal settings. For example, there is some reason to expect that the quality of feedback to an *S* may change systematically with time. In confounded feedback, the variable error in the feedback with respect to the response probably decreases with time (as each *S* "learns"), but constant errors stabilize. Individual learning experiments have generally tended to maintain such conditions constant over trials rather than to systematically vary them. An interesting exception occurs where a response is being gradually molded or differentiated. Here the R-S relationship is manipulated as a function of trials (or a joint function of trials and response adequacy) (11, chap. 8).

The two team measures, especially A_t, do not bear the same relationships with the feedback condition as does A_i. Indeed, differential trend effects are difficult to establish at all except through the use of liberal confidence levels. The difference in average value of A_t among the three feedback conditions are highly significant, however, and are worthy of comment because of an apparent difference in result as compared with A_i. Confounded feedback to a team member appears to result in at least as accurate, and possibly more accurate, team performance than direct feedback. This finding, if valid, points up an interesting phenomenon about distribution of feedback to teams where a group product rather than an individual product is the criterion. Since the task is one which permits compensation by one team member for difficult-to-extinguish constant errors of the second, dyads, *on the average*, would yield superior group products. The D_t findings attest to a greater difference in response magnitude between team members in confounded feedback than in direct feedback. This difference persists over trials and is consistent with the above hypothesis but does not obviate other explanations. Indeed, another tenable hypothesis is that the feedback values displayed to an *S* are much finer (more detailed) than he can discriminate kinesthetically. The inappropriate detail then acts as "noise" in the feedback producing response variability. In the confounded feedback condition, the feedback values (although as detailed as in direct feedback) are less variable from trial to trial than in direct feedback, since the former are based on an average of two response magnitudes. *S*s, in turn, may respond with less variability *about the goal score* from trial to trial. Further study of the phenomenon would be necessary both to replicate the differences in the A_t found in the present study and to examine intentionally the hypotheses stated here.

---

[5] Some preliminary attempts in this direction were made with these data, but the problems remain sufficiently horrendous to prevent adequate report of these at the present time.

The D_t measure, aside from corroborating the hypothesis above, has special interest for social psychologists. The response distribution in an interpersonal relationship receives considerable attention under such headings as role, role differentiation, role distribution, etc. (e.g., 6, 7, 10). The D_t measure is a considerable simplification of the usual role differentiation measures, but seems to represent the same type of measurement operation. The simplicity stems from the fact that a single response class (i.e., a single dimension of behavior) is used in this study. The analysis of the D_t measure under different feedback conditions points up a rather novel independent variable for social psychology. Response or role differentiation, D_t, bears the same simple relationship to the k_1/k_2 ratio as does A_i. The trend relationships are not as apparent, however.

The D_t measure also has some interesting implications for team reassemblies or team-member interchangeability. Conditions, whether they be feedback or other, which minimize the D_t value would appear to provide greatest flexibility for team reassemblies since teammates who respond similarly can be reassembled with a minimum of training. Teams requiring differentiation among a number of response classes for effective performance may require other feedback conditions.

SUMMARY

Knowledge of results (feedback) available to an individual working in a team may be a measure of: (*a*) his own response only, (*b*) a combination of his own and a teammate's response, (*c*) a teammate's response only. Each teammate may receive any of these three types of feedback.

The independent variable consisted of these three conditions of feedback in all six combinations. *S*s were used in pairs, and each pair was given 50 trials on a simple knob-turning task with one of these combinations of feedback conditions. A total of 27 pairs of *S*s was used in the study.

Three response measures were analyzed: (*a*) individual response accuracy, (*b*) team accuracy, (*c*) response (role) differentiation between team members. Analyses revealed significant differences between feedback conditions on each of these measures. Teams containing an *S* who received as feedback only his teammate's response showed poorest individual and team accuracy and greatest response differentiation. Differences among conditions in trend were consistently significant only for the individual accuracy measure. Learning rate on this measure was best when an *S* received information about his own response only and poorest when information consisted only of his teammate's response.

REFERENCES

1. ALEXANDER, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43,** 533–537.
2. ANNETT, J., & KAY, H. Skilled performance. *Occup. Psychol.*, 1956, **30,** 112–117.
3. BROWN, J. S. A proposed program of research on psychological feedback (knowledge of results) in performance of psychomotor tasks. In R. M. Gagne (Ed.), *Research planning conference on perceptual and motor skills. USAF, ATRC, HRRC Conference Rep.* 49-2. San Antonio, Tex. Lackland AFB, 1949.
4. GRAGG, D. B., & DOUGLASS, H. J. Conversion tables for selected airman classification battery scores and comparable scores on other selected service and civilian tests. *USAF Pers. Train. Res. Cent. Res. Rep.*, 1956, No. AFPTRC-TN-56-89.
5. GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53,** 141–154.
6. HARE, A. P., BORGATTA, E. F., & BALES, R. F. (Eds.) *Small groups.* New York: Alfred A. Knopf, 1955. Chap. 9.
7. HOMANS, G. C. *The human group.* New York: Harcourt, Brace, 1950.
8. LINDQUIST, E. F. Goodness of fit on trend curves and significance of trend differences. *Psychometrika*, 1947, **12,** 65–78.
9. MILLER, R. B. *Handbook on training and training equipment design. USAF WADC Tech. Rep.*, 1953, No. 53-136.
10. NEWCOMB, T. M., & HARTLEY, E. L. (Eds.) *Readings in social psychology.* New York: Henry Holt, 1947. Part VIII.
11. SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.

Received July 29, 1957.

A COMPARISON OF INTERVIEW AND OBSERVATION MEASURES OF MOTHER BEHAVIOR[1]

HENRIETTA T. SMITH

Vassar College

THE study of parent-child relationships has been largely dependent upon indirect measures of behavior, including such techniques as rating scales (1), questionnaires, interviews, and case histories. Merrill (4) found that controlled observations of interactions between mothers and children could be used to isolate and correlate significant aspects of behavior with a high degree of reliability. The present study is an attempt to evaluate the adequacy of interviews as compared with such controlled observations of actual behavior.

The interview method is economical for it enables the investigator to collect data on past and present maternal attitudes and behavior, and it allows comparable measures of both maternal and child behavior. But it introduces the major problem of the mother's ability, or desire, to present a true picture of her actual behavior. Errors may arise from involuntary misconceptions, faulty perceptions, or inaccurate recall.

Moreover, the mother may feel anxious or insecure about her normal behavior and may be unwilling to report it fully or without modifications that seem to her more acceptable to society. This is a more serious problem in establishing the validity of verbal reports as a measure of actual events because there are not sufficient cues to allow either discrimination of the mothers who consciously report inaccuracies or differentiation of the areas in which such errors occur.

Observation of behavior, when properly controlled, shows consistency of behavior and can give data on actual behavior under specific conditions. But it also involves the problems of the relationship between controlled conditions and the natural home environment and of the reliability of the sampling of maternal behavior. While both of these can be minimized by various techniques, they can be neither fully estimated nor controlled.

For the present research, dependency shown by the child at home and during the observation session was used as the constant child variable because earlier studies have shown certain relationships between specific maternal behavior and dependency (2, 3, 5, 6, 7). Maternal feelings about and behavior toward displays of dependency are the focus of the study. Replication of the findings of earlier studies will contribute to answering the two questions of primary concern here: (*a*) the relative adequacy and reliability of each method for studying maternal behavior; (*b*) the degree of similarity between reported and observed maternal behavior.

METHOD

Subjects

The *S*s were 30 mothers whose children, 18 girls and 12 boys, were enrolled in the Harvard Nursery School. The children's ages ranged from 3-0 to 4-1, with a mean age of 3-6. The mothers represented a highly selected group. Only four mothers had no college education; twenty-one had completed two or more years of college; five had completed one or more years of graduate work. It was expected, and found, that these mothers frequently would refer to publications on child rearing in order to explain their reasons for using a particular technique.

Procedure

The mothers were first observed for a 45-min. period while interacting with their children. Each mother was told that the *E* was interested in studying the child's play when she was present. She was reassured that no specific behavior was required and that she was free to do whatever she wished during the period. Both mother and child were then shown into the observation room, which was equipped with one-way vision windows, and the observer withdrew. After 30 minutes, the observer had the mother fill out a demographic questionnaire sheet in order to provide an experimental measure of her behavior toward the child's dependency solicitations when she was busy. As the observer withdrew again, the mother was told that there were 15 minutes remaining in the session.

At the end of the 45 minutes, *E* announced the end of this session and made arrangements for the interview. The interviews were conducted in the home by seven interviewers within two to four weeks after the observation session, and with only the mother present. Four interviews followed at longer intervals because of serious illness of mothers or children.

The training for the interviewers consisted of one month of intensive study of the interview schedule and the rating scales, as well as practice in actual interviewing. The practice interviews were criticized with reference to such factors as rapport with the mother, manner of interviewing, and adequacy of the answers when rated. The first few field interviews were also discussed with the interviewer. All interviews were recorded by means of an Audograph and scored directly from the records.

---

[1] This article is based on the author's Ph.D. dissertation completed at Radcliffe College in 1953 under the chairmanship of R. R. Sears, whose help is gratefully acknowledged.

Equipment

The observation room contained a large table on which were several current magazines of various types, an upholstered chair, a child's table with two chairs, and a toy chest which could also be used as a bench or table. The toys were: (*a*) a set of finger paints and paper; (*b*) a set of small tools with thumbtacks and pieces of wood; (*c*) a set of colored blocks; (*d*) a jigsaw puzzle of a map of the United States; (*e*) a blackboard with erasers and chalk; (*f*) a partially dressed doll with the rest of its clothes; and (*g*) a set of interlocking blocks for building.

Scoring

To record the observed actions, a unit of behavior was defined as the equivalent of a simple sentence, although one element might be implied. The behavior of both the mother and the child was recorded on a scoring sheet containing a column for each category. The units were set down in rows in order of occurrence. Thus it was possible to indicate the sequence of interactions as well as the number and nature of elements scored for each person.

The child's dependency behavior was considered in terms of his stimulus and response activities; i.e., those actions which seemed to be primarily directed toward stimulating the mother to give help or attention which she did not volunteer, and actions that were primarily in response to situations which the mother seemed to set up. It was important to differentiate the conditions under which he sought help, the methods he used to solicit aid, and the kind of help or attention he desired. The mother's behavior was also thought of in terms of her stimulus and response actions, the latter being defined as behavior that resulted directly from specific stimulation by the child and served to end an on-going sequence. Although these actions might in turn stimulate further activity, they were differentiated on theoretical grounds from such maternal stimulus acts as degree of control exerted, specificity of control, and degree of contact imposed by the mother.

This conceptualization of the stimulus and response aspects of both child and mother behavior was used in the development of the categories for scoring the observed behavior. Table 1 presents the categories used for scoring both mother and child behavior. A category of positive social interaction was used to record verbal interchanges that were not concerned with dependency interactions, and therefore was given lowest priority in scoring and was not used in the final analysis of data. The observation session was timed and the end of each 15 minutes indicated on the scoring sheet.

Reliability was established with one other observer. Category reliability for those used varied from .75 to 1.00. Observer reliability for five sessions scored independently by the two observers ranged between .88 and .96, with a mean reliability of .94.

TABLE 1

SCORING CATEGORIES FOR OBSERVED BEHAVIOR

1. X—behavior in which the child engages to evoke the attention or affection of the mother; e.g., whining, leaning against her
2. Asking verbal help—asking for generalized directions or assistance; seeking information rather than direct aid
3. Asking material help—asking for specific physical help
4. Asking reward—asking for direct praise, reassurance, affection, or attention
5. Teaching—giving the child new information which is related to, but not dictated by, the situation; e.g., screwdrivers can be used to tighten screws or to lift out thumbtacks
6. Structurizing—giving generalized information from which the child may select his own behavior. The mother sets up the situation for the child, but gives him autonomy within it
7. Giving positive directions—giving specific directions or demands which are precise requirements for the child to meet
8. Giving negative directions—giving specific directions or demands which prohibit the child's ongoing behavior, but which do not offer alternatives to it
9. Doing—physically carrying out of tasks by the mother for the child, usually preceded by some directions; an attempt to demonstrate to the child how the task should be done
10. 0—asking permission to help when the child has expressed no need for aid
11. Giving reward—giving attention, nurturance, praise, or affection at the end of an action sequence or in response to the child's request
12. Comply—performance of a task which has been presented in the form of a demand, or verbal compliance to such a stimulus as the child's asking permission
13. Noncomply—lack of response to requested performance; including ignoring the demand
14. Punishment—verbal reprimands, physical punishment, disagreement, or derogation; usually comes at the end of an action sequence or serves to end a specific sequence
15. Leaves field—either participant indulges in activity or apparent thought which is unrelated to the other person, or to the total situation
16. Total observed dependency—summation of child scores under the first four categories; computed at the end of each session

The interview consisted of 36 open-ended questions relating to such variables as infant care and training, present demands made upon the child, amount and kinds of attention requested by the child at home, and the ways in which the mother responded to dependent behavior in the child. The scales on which ratings of mother behavior were made were grouped into the classifications of factual information, such as separations from the child, usage of "do" and "don't," and offering of help; information on feelings of the mother for the child, such as degree of warmth, nature of affectional relationship, and rejection by the mother; and information about the restrictions imposed upon the child, such as scheduling or feeding, restrictiveness of physical mobility, and overprotectiveness defined as excessive protection of the child in a majority of situations where the child might be considered capable of independent action.[2]

Scales on child behavior were classified according to the nature of dependency solicitations (emotional, such as clinging or whining, or physical, such as wanting help while dressing), the conditions under which dependency occurs, and the areas in which the child strives for independence.

Interview scales which had a reliability of .60 or less were discarded and not used in analyzing the data. Since the remaining scales were divided into five points with 1 and 5 representing the extremes, and 3 the midpoint, two measures of rater reliability were used. The total rater reliability of the *E* and five other raters for nine interviews was .71 for complete agreement and .94 when one-point disagreements were included.

RESULTS

In comparing relative adequacy of these two methods, three measures were used: (*a*) consistency of observed behavior when a required task was interpolated; (*b*) replication of earlier findings on the antecedents of dependency behavior; and (*c*) similarity between observed behavior and reported behavior under similar conditions.

Table 2 presents the changes in behavior from the first 15 minutes of the session to the second 15 minutes, and from the first 15 minutes to the last. These are measured by the *t* test for the significance of differences between correlated means, and by product-moment correlations between the frequencies of each category for individual variability.

Both *structurizing* and *giving reward* decreased from the first to the second 15 minutes, possibly as a result of the child's increasing ease in the situation or of the mother's decreasing feeling of any need to stimulate the child to new and better activities.

From the first 15 minutes to the last 15 minutes, there was a significant decrease in *teaching* and *structurizing*, along with a significant increase in *give negative directions* and in *noncompliance*. Thus, when busy, the mothers became less rewarding and more restrictive. In general, however, there is evidence that individual mothers tended to be consistent in their behavior.

TABLE 2

CHANGES IN MOTHER BEHAVIOR DURING THREE PARTS OF OBSERVATION SESSION

(N = 30)

Categories	Mean Frequencies			t_{12}	t_{13}	r_{12}	r_{13}
	First 15 min.	Second 15 min.	Third 15 min.				
Teach	8.07	8.67	4.17	0.49	3.82***	.39*	.37*
Structurize	12.20	9.43	6.43	2.98***	6.14***	.64**	.36*
Give positive directions	23.53	24.27	22.07	0.43	0.85	.15	−.02
Give negative directions	3.10	3.83	4.23	1.43	2.02	.17	.38*
Give reward	13.67	10.23	15.23	2.92***	0.97	.30	.25
Comply	9.17	8.70	10.00	0.46	0.73	.48**	.21
Noncomply	3.11	3.67	14.67	0.07	7.71***	.72**	.21

* $p \leq .05$.
** $p \leq .01$.
*** $p \leq .001$.

Table 3 shows the correlations between five interview scales (selected because of their theoretical importance in the areas of warmth, frustration, and current restriction) and measures of dependency. Table 4 presents the correlations, in terms of *r*, between observed mother behavior and selected measures of dependency.

From Tables 3 and 4 the following significant relationships were found:

a. From interview data: *total dependency* is positively related to overprotectiveness; and *dependency* is positively related to rejection, especially in girls.

b. From observation data: *dependency* is negatively related to mother compliance and to mother giving reward, both of which are measures of lack of frustration; the more the mother *leaves the field* or *punishes* the child, the more frequently the child asks for material help, apparently to bring about an amicable close interaction; and the more the mother *complies* with the child's requests, the less verbal help or attention is requested.

c. From both observation and interview data: *dependency* in girls is positively related to warmth of the mother; and *dependency* is negatively related to amount of punishment given by the mother.

These replicate earlier results found on the relationship between mother behavior and dependency in children.

Since ratings on the techniques preferred by the mothers in their interviews followed the

---

[2] A 5-page table giving the interview questions, and a 15-page table giving the interview rating scales have been deposited with the American Documentation Institute. Order Document No. 5635 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting $2.00 for microfilm or $3.75 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 3

CORRELATIONS BETWEEN INTERVIEW ANTECEDENTS AND MEASURES OF DEPENDENCY

Interview Antecedents	Total Dependency on Interview	Total Dependency on Interview		Total Dependency on Observation	
		Boys	Girls	Boys	Girls
Schedule feeding	−.13	−.13	−.08	.71**	.04
Severity of weaning	.08	.13	.04	.14	.24
Rejection	.51**	.27	.65**	.10	.18
Affectional relations	.27	−.06	.47	−.01	.46*
Overprotection	.49**	−.03	.11	.03	.28

* Significant at .05 level.

** Significant at .01 level.

TABLE 4

CORRELATIONS BETWEEN OBSERVATION ANTECEDENTS AND MEASURES OF DEPENDENCY

Observation Antecedents	Total Dependency on Interview	Total Dependency on Observation	Ask Material Help	Ask Verbal Help
Comply	.18	−.36*	.09	−.36*
Noncomply	.07	.02	−.29	.22
Give reward	.07	−.36*	−.18	−.16
Leave field	−.20	.29	.51**	−.14
Punish	−.39*	.10	.40*	−.18

* Significant at .05 level.

** Significant at .01 level.

definitions set up for categories of observed behavior, reports of behavior when mothers were busy were compared with their behavior during the last 15 minutes of the observation session by means of the *t* test. No significant results were found. Using the same test for the entire observation session and reported techniques, only two significant differences were found. The mean observed use of structurizing was higher, at the .01 level, for those mothers who reported its use than for those mothers rated as not using it; giving positive directions was higher at the .05 level.

It was found that 70% of the mothers reported using techniques which they were also observed to use. Thus, by the sign test, ratings on techniques from the interview were positively related to the total observed techniques at the .05 level of significance. The *t* test revealed that mothers reported more accurately their use of "don't" than of "do." The mean observed use of negative directions throughout the session was higher, at the .05 level, for mothers who reported high usage of "don't" than for those who reported high usage of "do." In contrast, the mean observed use of positive directions was not significantly different for those mothers who reported high usage of "do" than for those who reported high usage of "don't." This result seems to indicate that mothers were more conscious of actions that they considered "bad" and were therefore able to make better estimates of how frequently they occurred.

DISCUSSION

The general consistency of behavior during the first half-hour of observation seems to indicate that the mothers were giving a reliable sampling of their normal behavior under similar conditions at home. When this is contrasted with the increased individual variability of behavior during the last 15 minutes, it can be hypothesized that the introduction of the task not only changed the mothers' activities, but also changed their set within the situation. Thus, mothers who approached the situation with a definite set as to how they should, or would, act throughout the 45 minutes, could not maintain this behavior when the external conditions were changed. But this alone does not invalidate the reliability of the method.

To supplement this measure, it was assumed that the method which best replicated previous findings on the relationship between mother behavior and dependency in the child would be the more adequate. Since the Iowa study (5) replicated findings of earlier studies, it was used as the criterion in this area. Mother behavior from the interview data seems to contribute more of the significant findings that replicate the Iowa study. But when *t*'s were computed on means of mother and child behavior during the last 15 minutes, it was found that increase in mother *noncomply* was positively related to increase in child *ask reward* at the .05 level, and mother *give reward* was negatively related to whining behavior at the .02 level. This relationship between warmth and whining is also similar to the Iowa report's finding on negative attention. Thus, the general results of this study replicated earlier findings on dependency, but neither method contributed more than the other.

A basic assumption for the study was that the observed behavior represented a criterion against which reported behavior could be validated. To test this, the mean and *SD* for changes in each observation category from the first to the last 15 minutes of the session were determined. The mothers whose scores on each category lay more than one *SD* from the mean were identified. Most of the mothers' scores lay outside this range for one or another category, but only six mothers showed such a deviation on 40% or more of the categories.

The introduction of a new task was constant; therefore, it was reasoned that those mothers who showed a change of behavior greater than that shown by four-fifths of the mothers were probably giving a less representative sample of their usual behavior during the first half-hour of observation than were the majority. Whether these six differed because of the degree and kind of motivation under which they took part in the study, or because they were more anxious and thus more defensive about their normal behavior was explored by testing the hypothesis that their answers on the interview would show higher conformity to a group mean, especially for questions which might be anxiety-arousing, than the other mothers' answers.

Using the *t* test, this group was compared with the 24 other mothers on 10 interview questions, 5 of which were judged to involve some anxiety and 5, selected at random, which were considered to be neutral. There were no significant differences found. When the means and *SD*s for these 10 scales were computed, it was found that the 6 mothers were consistently within one sigma of the mean; of the other 24 cases, from 8 to 12 mothers on each scale fell outside this range. It was not, however, a consistent group who did so. This seems supportive evidence that "defensive" mothers can, and do, give interview information which may be misleading in its very normality. In addition, they may not be discovered during a brief observation under constant conditions.

From the preceding results, there is little evidence for deciding that either interviews or controlled observations are better for securing valid measures of actual mother behavior. Since observed behavior is consistent for most mothers; since both measures yield results that replicate previous findings on the antecedents of dependency; and since most mothers report similar behavior to that observed, the interview seems to be the preferable method because it permits study of a wider range of behavior in a shorter time than can an observation session. It is important to remember, however, that "defensive" mothers can be more easily discriminated by the variability of their behavior under observation conditions than by their answers to interview questions.

SUMMARY

In order to compare the adequacy of interviews and observations as methods, 30 mothers were studied by means of individual interviews and controlled observation while interacting with their children. Neither method was found to be completely superior to the other by the measures used: consistency of observed behavior when a required task was interpolated, replication of earlier findings on the antecedents of dependency behavior, and similarity between observed and reported behavior. In general, it was concluded that the interview was a preferable method only because it allows coverage of a wider range of behavior than does an observation. However, it was found that observation of behavior discriminated "defensive" mothers better than did the interview.

REFERENCES

1. CHAMPNEY, H. The variables of parent behavior. *J. abnorm. soc. Psychol.*, 1941, **36**, 525–542.
2. GEWIRTZ, J. L. Succorance in young children. Unpublished doctoral dissertation, State Univer. of Iowa, 1948.
3. LEVY, D. M. *Maternal overprotection.* New York: Columbia Univer. Press, 1943.
4. MERRILL, BARBARA. A measurement of mother-child interaction. *J. abnorm. soc. Psychol.*, 1946, **41**, 37–49.
5. SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, P. S. Child-rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, **47**, 135–236.
6. SMITH, H. T. A comparison of interview and observation measures of mother behavior. Unpublished doctoral dissertation, Radcliffe College, 1953.
7. SYMONDS, P. W. *The psychology of parent-child relationships.* New York: D. Appleton-Century, 1939.

Received July 29, 1957.

NEURAL SATIATION, REACTIVE INHIBITION, AND INTROVERSION-EXTRAVERSION[1]

ALLAN RECHTSCHAFFEN[2]

Northwestern University

EXPANDING upon earlier theorizing by Pavlov (19), Eysenck (6) has hypothesized that extraverts develop more reactive inhibition (I_R) than introverts. Eysenck further noted similarities between I_R and neural satiation, the construct proposed by Köhler and Wallach (18) to underlie figural aftereffects. The similarities of the constructs and the phenomena from which they were inferred were also noted by Köhler and Fishback (17) and Duncan (4). On the basis of such similarities, Eysenck further proposed that extraverts are more susceptible to the development of neural satiation than introverts.

In his first test of these hypotheses, Eysenck (6) compared the kinesthetic aftereffects of introverts and extraverts. The introvert and extravert groups were defined by the dual criteria of cutoff scores on Guilford's R (Rhathymia) Scale (8) and psychiatric diagnoses; obsessionals, reactive depressives, and anxiety neurotics were included in the introvert group, while hysterics and psychopaths were included in the extravert group. The results were interpreted as showing significantly greater aftereffects for extraverts, hence supporting the hypothesis. In a more recent study, Eysenck (5) correlated reminiscence scores (which he proposed as a measure of I_R) on a pursuit rotor task with the extraversion-intraversion scores of university students on the Maudsley Personality Inventory, again obtaining results that he interpreted as confirming his hypothesis. Eysenck has also cited the work of Franks (7) which showed that introverts developed conditioned responses more readily than extraverts as providing further support for his theories. While some of the above findings certainly appear suggestive, certain procedural and statistical aspects of these investigations cast some doubt on Eysenck's interpretation. These matters are discussed at a later point.

The present study was designed to investigate further the hypotheses cited above. Three hypotheses were investigated: 1. Extraverts develop greater visual aftereffects than introverts. This hypothesis follows directly from Eysenck's theory, if one can assume that satiability is a fairly general trait extending across modalities. 2. Extraverts develop more I_R on a motor-learning task than introverts. In the present investigation, inverted alphabet printing was used to measure I_R. This choice of task seems appropriate in that several investigators (e.g., 2, 20) have demonstrated large amounts of I_R to build up with massed practice in inverted alphabet printing. 3. Amounts of aftereffect and I_R exhibited by individuals are positively correlated. This hypothesis follows from the thinking of Eysenck, Köhler and Fishback, and Duncan regarding the similarities of the two phenomena. In his summary of the literature on this topic, Duncan has pointed out that not only are I_R and aftereffects grossly similar in that they both represent response alteration following continued response evocation, but that they also have strikingly similar accumulation and decay functions. Such similarities lead to the expectation that individual differences in I_R would be correlated with differences in aftereffect, assuming either that the same neural substrate is involved in both phenomena or that individual susceptibility to the phenomena is relatively independent of the particular neural area being stimulated.

METHOD

The over-all plan of the investigation involved obtaining from each *S* a measure of introversion-extraversion, I_R, and visual aftereffect.

Subjects and Measurement of Introversion-Extraversion

Guilford's R Scale, which had been administered in group form to an entire introductory psychology class, was used to define the introversion-extraversion dimension, high R scores representing extraversion. From the class, 96 *S*s were obtained as volunteers, a few *S*s being

---

[1] This paper is based upon a doctoral dissertation submitted to Northwestern University. The author is grateful to Janet A. Taylor for her guidance in this research.

[2] Now at the University of Chicago Medical School.

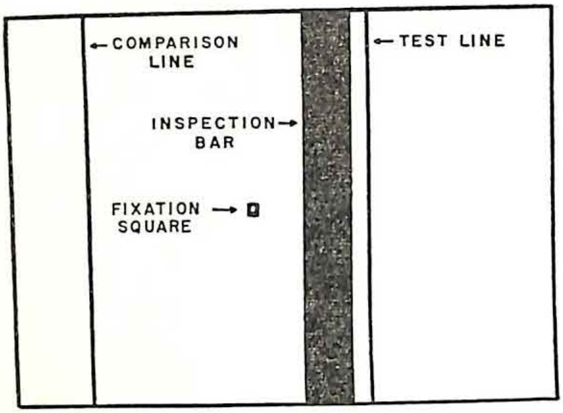

FIG. 1. RELATIONSHIP OF TEST AND INSPECTION FIGURES IN THE VIEWING AREA

(Figures are superimposed so that their fixation squares coincide)

requested to participate in order to obtain an equal number with R scores falling above and below the class mean and an equal number of males and females. The mean R score of the 96 *S*s was 38.8; the *SD* was 11.5; the range was 8–70; the distribution was fairly symmetrical, tending towards bimodality.

Procedure

*S*s were tested individually on the inverted alphabet-printing and visual aftereffect tasks during a single experimental session lasting approximately 45 min. The order of task presentation was counterbalanced, half of the *S*s being given the alphabet printing first and half the visual aftereffect task first. Within each of these groups, the same counterbalancing was maintained with regard to sex and position above or below the sample mean on R scores. For all *S*s, a rest of 5 min. separated the administration of the two tasks.

Inverted Alphabet Printing

*S*s were told that the inverted alphabet-printing task was designed to measure a facet of their intelligence. They were given standard instructions for the task (e.g., 2), including a demonstration that seven letters appeared the same inverted as they did right side up. *S*s worked in 20-page booklets, each page containing a 10 × 6 matrix of ½-in. squares.

*S*s were first given five 30-sec. trials separated by 30-sec. rest intervals (distributed practice). At the end of each 30-sec. practice interval, *S*s were instructed to turn to the next page of the test booklet, sit quietly for the 30-sec. rest interval, and to start on the new page with the next consecutive letter (following the last one they had written on the preceding page) when *E* gave the signal.

Following the distributed practice, *S*s were given a 1-min. rest interval and then 10 consecutive 30-sec. trials (5 min. of massed practice). At the end of each 30-sec. trial, *E* signaled to *S* to turn to the next page of the test booklet and proceed immediately with the next consecutive letter.

After 49 *S*s had been run under the conditions described above, it was decided to include a reminiscence measure in the study. Hence, the remaining 47 *S*s were given a 1-min. rest following the last trial of massed practice and two consecutive 30-sec. trials following the rest. Of the 47 *S*s thus tested, 25 *S*s had R scores of 39 or above, and 22 *S*s had R scores of 38 or below.

Visual Aftereffect

A figure (see Fig. 1) similar to one employed by Wertheimer (21) was used.[3] During inspection, only the fixation square and inspection bar were shown, theoretically satiating the region corresponding to the visual space to the right of the fixation square. When the test figure (fixation square, test line, and comparison line) were subsequently presented, the satiation should have become manifest in an apparent displacement of the test line away from the fixation square. By having *S*s adjust the comparison line to appear the same distance from the fixation square as the test line, these apparent displacements could be measured as differences between preinspection and postinspection settings of the comparison line.

An apparatus similar in principle to one employed by Hammer (9) was used. The essential components were two clear-glass plates firmly set in a slotted frame. On one plate were fixed the fixation square and inspection bar of the inspection figure. By appropriate movement of this plate in the slotted frame, the test and inspection figures could be alternately exposed in the opening of a reduction screen. On the other glass plate, which lay proximal to *S* and almost immediately next to the first one, was fixed the comparison line. By turning an adjustment wheel which was attached to a pulley system on the frame, *S*s could move the plate containing the comparison line either left or right. A hairline indicator attached to the comparison-line plate moved against a millimeter rule located at the rear of the apparatus, thus enabling *E* to make readings of *S*s' settings of the comparison line. A transluscent milk-glass plate was situated behind (distal to *S*) the two plates cited above. The viewing area was illuminated from behind this transluscent plate, providing a homogeneous field, except for the test and inspection figures, which was free from distinguishing cues.

After being instructed in setting the comparison line the same distance from the fixation square as the test line, *S*s were given three practice trials of four adjustments each. (The terminology of referring to a set of four adjustments as a trial will be maintained throughout this report.) *S*s were trained to make the four adjustments within a trial in rapid succession, and most *S*s learned to make four adjustments in about 30 sec. The first and third adjustments of a trial were always ascending, i.e., the comparison line was set too close to the fixation square by *E* and had to be moved further away by *S*. The second and fourth adjustments were descending. The exact presettings of the comparison line were varied by *E* in an unsystematic manner. On all adjustments, *S*s were allowed to move the comparison line in either direction in order to make finer adjustments.

---

[3] A more exact description of the figure and the apparatus used, including all the dimensions involved in each, can be obtained from the original dissertation which is available from University Microfilms, Mic 56-5336, for $1.50.

Following practice trials, *S*s were given a 2-min. rest during which they were given instructions for the experiment proper. The instructions strongly stressed the importance of maintaining constant, faithful fixation on the fixation square during inspection and test periods. Then the following schedule was maintained: *S*s were given a preinspection trial, a 1-min. rest, 40 sec. of fixation on the inspection figure, and a postinspection trial. Two additional determinations of aftereffect were made for each *S*, making a total of three sequences of preinspection trial, rest, inspection, and postinspection trial. A rest of 3 min. was allowed between each sequence.

All adjustments of the comparison line were made monocularly with the right eye, while the inspection figure was always fixated with the left eye.[4] This procedure was introduced, because, during pilot work in which inspection and test were accomplished with the same eye, *S*s frequently reported strong negative afterimages of the inspection figure in postinspection trials. These afterimages usually caused *S*s much difficulty in maintaining fixation and making adjustments.

RESULTS

Visual Aftereffect

For all 96 *S*s, the mean amount of aftereffect (differences between preinspection and postinspection trials) was in the expected direction for all three inspection periods. Each of these means (4.1, 4.0, and 4.2 mm. respectively) was significantly different from zero ($p < .01$). Differences between the three preinspection trials were negligible and not significant, indicating that there was probably fairly complete dissipation of aftereffect during the 3-min. rest interval following the first and the second postinspection trials. This consideration enables us to view the three determinations of aftereffect as independent measures for purposes of estimating reliabilities. A single aftereffect score was determined for each *S* by summing his aftereffects for the three inspection periods. The reliability of these scores as estimated by Hoyt's r (10) was .52. Intercorrelations of the three independent measures were also computed. These were .25 for Trials 1 and 2, .25 for Trials 1 and 3, and .27 for Trials 2 and 3 ($p < .02$ in each case).

Thus the task did produce systematic, significant aftereffects as would be predicted by satiation theory, and significant, albeit low, reliability was found among the three measures. However, the results of the study do not support the hypothesis that there is a relationship between amount of visual aftereffect and R scores, our personality measure. The r between the aftereffect scores and the R scores for all 96 *S*s was .14. While the relationship was in the predicted direction, the magnitude of the correlation fell short of the .20 required for significance at the .05 level of confidence with 94 *df*.

In view of the possibility that only extreme introversion and extraversion are related to amount of aftereffect, an additional analysis was undertaken. Two subgroups were formed from the 96 *S*s. These groups were the "extreme extraverts," which consisted of the 35 *S*s with the highest R scores (44 or above) and the "extreme introverts," which consisted of the 35 *S*s with the lowest R scores (32 or below). The aftereffects of the 35 extreme extraverts were compared with the aftereffects of the 35 extreme introverts. As can be seen graphically in Fig. 2, the extreme extraverts appear to have somewhat larger aftereffects. However, the difference was not statistically significant. The extreme extraverts had a mean aftereffect score of 13.6 mm. as compared to 10.3 mm. for the extreme introverts, the t value for the difference being only 1.07.

Inspection of Fig. 2 also reveals in each group the expected development of aftereffect with inspection and dissipation with rest. It will be noted that the extreme extraverts tend to have lower preinspection settings than the extreme introverts. In none of the three trials, however, did the preinspection settings of the two groups differ significantly, nor were preinspection settings for all 96 *S*s significantly correlated with amount of aftereffect.

Inverted Alphabet Printing

For an I_R measure, an "inhibition score" was calculated for each *S* by subtracting the number of letters printed (correctly or incorrectly) during the last five trials of massed

---

[4] It is possible that the magnitude of visual aftereffects is reduced by using different eyes for inspection and test periods, but the author knows of no systematic studies on the subject. One possible interpretation of the bilateral transfer of satiation effects in the present study is that since there is bilateral representation of the macula in the occipital lobes, since the macula subtends a visual angle of approximately 3°, and since the inspection bar was included within 1°40′ of visual angle from the center of the fixation square, inspection could have produced satiation in corresponding portions of both occipital lobes.

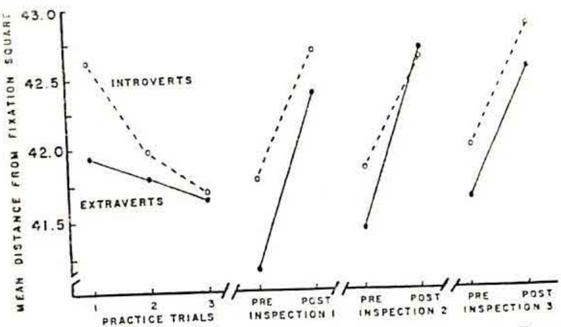

FIG. 2. MEAN DISTANCE PER ADJUSTMENT OF COMPARISON LINE FROM INSPECTION SQUARE IN MM. FOR PRACTICE, PREINSPECTION, AND POSTINSPECTION TRIALS, COMPARING 35 EXTREME EXTRAVERTS AND 35 EXTREME INTROVERTS. AMOUNT OF AFTEREFFECT IS SHOWN AS THE DIFFERENCE BETWEEN PREINSPECTION AND POSTINSPECTION TRIALS

practice from the number of letters printed during the five trials of distributed practice. Thus the higher the score the higher the I_R, e.g., a greater score during the early, distributed practice than during the later massed practice would indicate more I_R than a poorer score during distributed practice.

Using the inhibition scores thus derived, the results of the study do not support the hypothesis that there is a relationship between the accumulation of I_R and R scores. The *r* between the inhibition scores and the R scores for all 96 *S*s was .18.[5] While the relationship was in the predicted direction, it was not statistically significant.

A comparison was also made between the inhibition scores of the 35 extreme extraverts and 35 extreme introverts (see Fig. 3). It is readily apparent that although the performance curve of the extreme extraverts tends to remain above that of the extreme introverts, the slopes of the curves remain essentially similar throughout, contrary to the divergence expected if more I_R were being built up in one group than another. This finding would suggest that there is no substantial relationship between R score and accumulation of I_R. The extreme extraverts had a mean inhibition score of -10.9 as compared to a mean inhibition score of -15.7 for the extreme introverts. The difference was in the predicted direction but was not significant ($t = 1.26$). An analysis of variance for all 15 trials showed no significant difference between the groups on level of performance and no significant interaction between groups and trials.

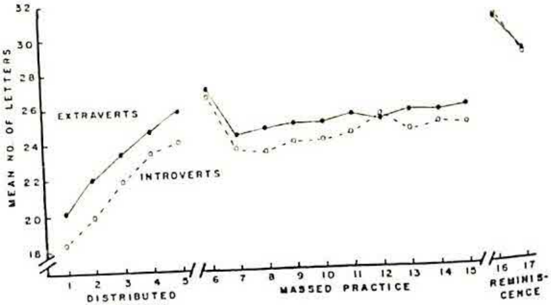

FIG. 3. COMPARISON OF PERFORMANCE OF 35 EXTREME EXTRAVERTS AND 35 EXTREME INTROVERTS ON INVERTED ALPHABET PRINTING FOR TRIALS 1–15[6]

It should be noted that the inhibition measure discussed above differs from that used by Eysenck (5), which was reminiscence. Since reminiscence represents a recovery from inhibition rather than inhibition per se, the use of the reminiscence score as an index of I_R appears to rest on the assumption that amount of recovery is positively related to the amount of I_R accumulated. While such an assumption seems reasonable, Eysenck (6) had postulated that not only do extraverts develop more I_R, but also that they dissipate it more slowly. Thus conflicting predictions result from Eysenck's theory when reminiscence scores are used. On the one hand, it seems reasonable to assume that if extraverts develop more I_R, they should show more recovery, i.e., a greater difference between performance at the end of massed trials and performance on reminiscence trials. On the other hand, since extraverts theoretically recover more slowly from I_R, it might be expected that they would show less reminiscence. Possibly, the two factors could both be operating and mask each other.

In spite of the above objection, a reminiscence score was obtained for 47 *S*s to offer some basis for comparison with Eysenck's study. Reminiscence scores were obtained by subtracting the number of letters printed during the last two trials of massed practice from the number of letters printed during the reminiscence trials.

---

[5] The *r* of .18 between inhibition scores and R scores is significant at the .05 level of confidence with 94 *df* when a *one-tailed* test is used as the criterion of significance. However, it should be noted that the occurrence of one such *r*, significant at the .05 level with a one-tailed test, from among the four *r*'s calculated in testing the major hypotheses, is not very unlikely on the basis of chance alone.

[6] The reminiscence trials are for 20 extraverts and 17 introverts; they should not be used for comparing extraverts and introverts on absolute amount of reminiscence, since the preceding massed practice curves were based on a larger *N*.

The *r* between inhibition and reminiscence scores for the 47 *S*s receiving both measures was .35, which is significant at the .02 level of confidence. While the above correlation is probably somewhat spurious by reason of the fact that the number of letters printed on the last two trials of massed practice enters into both the inhibition and reminiscence scores, the finding tends to support the assumption that the more I_R accumulated, the greater the reminiscence. However, as was pointed out above, differential rates of recovery for extraverts and introverts may be confounded in the reminiscence measure with amount of I_R accumulated.

The *r* between the reminiscence scores and the R scores for the 47 *S*s who received reminiscence trials was .08. With 45 *df*, an *r* of .29 is required for significance at the .05 level of confidence. Of the 35 extreme extraverts, 20 received reminiscence trials, and of the 35 extreme introverts 17 received reminiscence trials. These 20 extreme extraverts and 17 extreme introverts achieved mean reminiscence scores of 8.8 and 8.6 respectively, the difference between these means yielding a *t* value of only .06.

Relationship Between Visual Aftereffect and I_R

Although the magnitude of neither I_R nor aftereffects was significantly related to introversion-extraversion scores, it is still possible that the magnitude of aftereffects is correlated with the magnitude of I_R as hypothesized earlier. The *r* between the aftereffect and inhibition scores for the 96 *S*s was, however, only .04. The *r* between the aftereffect and reminiscence scores was only −.02.[7]

DISCUSSION

The results do not support the hypotheses that amount of aftereffect is related to introversion-extraversion, that accumulation of I_R is related to introversion-extraversion, or that the amount of aftereffect exhibited by an individual is related to his accumulation of I_R. While the results were in the expected direction in each case, the degree of relationship demonstrated was uniformly small, and in no case did it reach statistical significance.

There are several differences between the present study and those of Eysenck which may be partly responsible for the apparent divergence in results, e.g., different *S*s, differences in modalities being tested, task differences, and differences in experimental procedure. Since none of these differences appear to violate any of the boundary conditions specified by Eysenck for testing his hypotheses, however, it is clear his theories concerning the relationship of introversion-extraversion, I_R, and aftereffect are not supported under the experimental conditions described in this study. And, as previously mentioned, there are grounds for uncertainty concerning Eysenck's findings themselves.

Consider first Eysenck's study (6) relating kinesthetic aftereffects to introversion-extraversion. It will be remembered that the aftereffect scores in the present study were obtained from the differences between preinspection and postinspection settings of the comparison line on each trial. Eysenck's procedure was different. He obtained a single preinspection setting (an average of four adjustments) and four postinspection settings, one after each of four successive inspection periods. As aftereffect scores for each of the four trials, Eysenck used the per cent change from the original preinspection setting. Since he was also interested in relating rate of dissipation of aftereffect to personality, Eysenck also obtained settings after 5 min. of rest and again after an additional 10 min. of rest. The major evidence offered by Eysenck for the difference in aftereffect between introverts and extraverts was a Hotelling's *T* test which invalidated the null hypothesis at between the .01 and .05 levels of confidence. The calculation of Hotelling's *T*, an over-all test of significance, was based upon nine separate aftereffect scores which were as follows:

---

[7] Since the original writing of this paper, the author has learned of a study (Meier, M. J. Interrelationships among personality variables, kinesthetic figural aftereffects and reminiscence in motor learning. Unpublished doctoral dissertation. Univer. of Wisconsin, 1956) which showed a significant *negative* correlation between amount of reminiscence in inverted alphabet printing and amount of kinesthetic aftereffect.

1. Average per cent change after 30 sec. of inspection; 2. average per cent change after 60 sec. of inspection; 3. average per cent change after 90 sec. of inspection; 4. average per cent change after 120 sec. of inspection; 5. sum of the above four scores; 6. maximum single per cent change obtained from each *S*; 7. average per cent change after 5 min. of rest; 8. average per cent change after an additional 10 min. of rest; 9. sum of the above two scores.

First, it appears likely that Scores 1–4 are not independent measures of aftereffect. In discussing his experimental procedure, Eysenck gave no indication that there were any rest pauses separating the four inspection periods, aside from the time taken to make the postinspection settings. Since aftereffects following comparable periods of inspection are known to last up to 90 sec. after inspection (9) and have been found to cumulate (21), we might expect that there would be an accumulation of undissipated aftereffect from trial to trial in Eysenck's procedure. While the amount of such accumulation could not be estimated from Eysenck's results (he used progressively longer inspection periods in the four successive trials, and accumulation of undissipated satiation is confounded with increased inspection time), it seems likely that some accumulation resulted and that the successive trials in his study cannot be considered to be independent measures of aftereffect. In addition, there is an obvious lack of independence among the first four scores and Scores 5 and 6, as well as among Scores 7, 8, and 9. It would appear to be a questionable procedure to use for an over-all test of significance such a series of nonindependent, spuriously intercorrelated scores, some of which do little more than superfluously repeat original trends.

Eysenck also computed *t* tests between introverts and extraverts for the nine scores considered individually. Of the nine *t*'s, one (for the 30-sec. inspection period) was barely significant at the .05 level of confidence using a one-tailed test. The other eight scores were significant at approximately the .10 level only, using a one-tailed test. While Eysenck's findings are suggestive, they are far from striking, and the failure to replicate his results in the present study is not particularly surprising.

A final question arises as to the relative adequacy of the aftereffect measures. The estimated reliability of the aftereffect scores in the present study was .52 as compared to estimated reliabilities of .78 and above reported by Eysenck for his measure of kinesthetic aftereffect. At face value, the measure used in the present study seems relatively poor for singling out a highly stable individual characteristic of satiability. Closer examination, however, discloses important differences in the derivation of the respective scores which makes the apparent relative unreliability of our own scores open to considerable question.

First, it will be noted that Eysenck's reliability coefficients were obtained by Hoyt's *r* which depends, in essence, upon the correlations among the several trials. As noted previously, however, the four test trials in Eysenck's measure of kinesthetic aftereffect are probably nonindependent by reason of the accumulation of undissipated aftereffect from trial to trial. For purposes of reliability estimation, this type of nonindependence must be considered as producing spuriously high correlations. Further, it will be recalled that Eysenck used a single preinspection setting as a base against which to calculate per cent change for all four postinspection trials. Insofar as the single preinspection setting enters into all four aftereffect measures, and insofar as we can expect interindividual variation in preinspection settings, the reliability of aftereffect is confounded with the constancy of the preinspection setting, once more producing a spuriously high reliability coefficient.

It is entirely likely that kinesthetic aftereffects are more reliable than visual aftereffects, because they are not so subject to the vagaries of fluctuating fixation. However, in view of the errors involved in Eysenck's estimation of reliability, this has not yet been adequately demonstrated. Certainly, the differences in reliability are not so great as uncritical evaluation of the reported reliabilities would lead us to believe.

Wertheimer (22) has reported a retest reliability of .82 for visual aftereffects. However, his measure of aftereffect was based on postinspection settings alone, rather than on differences between preinspection and postinspection settings. Inasmuch as we can expect interindividual variation in constant error (apart from the development of aftereffect), and inasmuch as we can expect some reliability of constant error in successive tests, Wertheimer confounds reliability of aftereffect with the reliability of constant error. For example, using Wertheimer's procedure with our own data, the estimated reliability was raised from .52 to .92

Wertheimer himself has more recently called attention to the importance of considering constant errors in the measurement of aftereffects (23).

Now we turn to the evidence Eysenck has presented as demonstrating the hypothesized relationship between introversion-extraversion and I_R. Prior to performing his own experiment on this particular hypothesis, Eysenck (6) cited the work of Franks (7) as supporting his theory. Franks found that introverts (dysthymics) showed a higher level of performance in eyelid conditioning than extraverts (hysterics), and Eysenck held this difference to result from greater I_R for the extraverts. There is reason to question this interpretation. According to the Hullian definition (11), which Eysenck has followed, I_R is produced by response evocation, the magnitude being dependent on the amount of work involved. Further, I_R cumulates under conditions of massed practice and becomes manifest in performance decrement (when groups receiving massed practice are compared with groups receiving distributed practice). Franks presented no indication of the intertrial interval employed in his conditioning procedure nor of the amount of work involved in each response, although it should be noted that the typical intertrial intervals and the amount of work involved in eyelid conditioning are not favorable for the accumulation of I_R. Had the conditions produced accumulation of I_R, however, and had the differences between extraverts and introverts derived from differences in I_R, then it would be expected that the differences in performance between the groups would increase as conditioning progressed, the amount of I_R cumulated becoming relatively greater for the extraverts. Inspection of the acquisition curves obtained by Franks shows this not to be the case. The two curves differ on the initial conditioning trials and show no systematic tendency to diverge on subsequent trials, an apparent contradiction of the I_R hypothesis.

In his own experiment relating introversion-extraversion to I_R, Eysenck (5) employed the pursuit-rotor task, giving three 5-min. periods of massed practice separated by two 10-min. rest periods. As was discussed previously, Eysenck used reminiscence (following the two rest periods) as a measure of I_R. It is important to note that Eysenck used the difference between the last three trials of prerest practice and the first trial of postrest practice as his reminiscence measure. This procedure was in contrast to the measure advocated by Ammons (1) who argued that a correction be made for warm-up (the rapid initial rise in performance following the first post-rest trial). Ammons reasoned that reminiscence should be measured as the difference between the last trial of pre-rest practice and the asymptote of the warm-up increment, so that warm-up effects might not mask the full amount of reminiscence occurring. Eysenck, however, has followed the argument of Denny, Frisby, and Weaver (3) that what Ammons has called "warm-up" is actually the extinction of $_SI_R$. By this reasoning, the use of the reminiscence measure suggested by Ammons would confound the effects of I_R and $_SI_R$.

Eysenck's results showed an *r* of .29 between the first reminiscence score (following the first rest period) and an introversion-extraversion measure. For his *N* of 50 *S*s, this was significant at the .02 level of confidence using a one-tailed test. The *r* between the second reminiscence score and the introversion-extraversion measure was only .10; that between both reminiscence scores and introversion-extraversion was .22; neither value is significant at the .05 level of confidence. Eysenck interpreted these results as showing a significant relationship between I_R and the personality measure. He explained that a lower correlation between introversion-extraversion and the second reminiscence measure was to be expected. Citing earlier findings by Kimble (12, 13, 14, 15, 16), Eysenck stated that $_SI_R$ can be expected to develop late in practice, and reasoned further that this increased amount of $_SI_R$ at the time of the second reminiscence measure probably increased error variance so as to make the second reminiscence measure a less reliable index of I_R than the first. Eysenck's case would appear to depend on the following conditions: there would have to be a fairly substantial difference in error variance between the two reminiscence situations to result in a shift from .29 to the .10 correlation; this in turn would have to result from very substantial differences in $_SI_R$ between the two reminiscence situations. Now Eysenck has suggested that the initial rise in performance following rest is the result of extinction of $_{S}I_{R}$. On this basis, it would be expected that the "$_{S}I_{R}$ extinction increment" following the second rest period would be substantially greater than that following the first rest period (assuming that amount of $_{S}I_{R}$ extinction is positively related to amount of $_{S}I_{R}$ accumulated during prerest practice). Inspection of the graph of performance presented by Eysenck, however, shows this not to be the case. The second "$_{S}I_{R}$ extinction increment" is only barely greater than the first. In view of this internal inconsistency in Eysenck's interpretation and data, and in view of dangers generally inherent in selecting significant findings as more valid than insignificant findings on a speculative basis, we are once more led to conclude that Eysenck's findings do not support his hypotheses conclusively.

While in general it may be said that all studies thus far performed on the relationship between introversion-extraversion and aftereffects, and between introversion-extraversion and I_R, have yielded results in the direction predicted by Eysenck's theory, in no case has the degree of relationship been shown to be very strong, and rarely have the findings been statistically significant. Acceptance of the theory seems to await more conclusive supporting evidence.

Leaving the personality variable and considering the nonsupportive evidence concerning the hypothesized similarity of I_R and neural satiation, one must be cautious regarding the possible interpretations. Certainly, the results of the present study would indicate that the two phenomena, as they are measured here, are not correlated. However, this does not mean that the phenomena are dissimilar; on a descriptive level there is no denying their similarity. Furthermore, it is even possible that I_R and neural satiation are based on the same neurological processes. If there is no great intrasubject consistency in the satiability of different neural areas, then the negative results of the present investigation could derive from the fact that different neural areas were being tapped in the measurement of the two phenomena. For the same reason, it might not be surprising to find that separate measures of I_R determined on different tasks were not correlated. Response correlation is useful as supportive evidence for similarities between phenomena, but the method is by no means decisive for ruling out similarities in the underlying processes.

SUMMARY

Expanding upon earlier theorizing by Pavlov, Eysenck has hypothesized that extraverts would show greater figural aftereffects (or neural satiation) and more reactive inhibition (I_R) on a motor task than introverts. Several investigators had noted similarities between neural satiation and I_R and had hypothesized that the neural mechanisms underlying both phenomena might be the same.

These various hypotheses were investigated in the present study by testing 96 *S*s on introversion-extraversion, visual aftereffects, and I_R. Introversion-extraversion was defined in terms of scores on Guilford's Rhathymia Scale. Visual aftereffects were measured with a specially constructed apparatus, and two measures of I_R were obtained from an inverted alphabet-printing task.

The results showed that the introversion-extraversion (Rhathymia) scores were not significantly correlated with either amount of visual aftereffect or I_R measures. Further, amount of visual aftereffect was not significantly correlated with the I_R measures. An additional analysis was made comparing the 35 *S*s with the highest Rhathymia scores (extraverts) with the 35 *S*s with the lowest Rhathymia scores (introverts). The two groups also did not differ significantly on the aftereffect and I_R measures.

The results were interpreted as failing to support the hypotheses concerning the relatedness of the three variables, and the findings were compared to those of previous investigations.

REFERENCES

1. AMMONS, R. B. Acquisition of motor skills: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
2. ARCHER, E. J. Postrest performance in motor learning as a function of prerest degree of distribution of practice. *J. exp. Psychol.*, 1954, **47**, 47–51.
3. DENNY, M. R., FRISBY, N., & WEAVER, J. Rotary pursuit performance under alternate conditions of distributed and massed practice. *J. exp. Psychol.*, 1955, **49**, 48–54.
4. DUNCAN, C. P. On the similarity between reactive inhibition and neural satiation. *Amer. J. Psychol.*, 1956, **69**, 227–235.

5. EYSENCK, H. J. Reminiscence, drive, and personality theory. *J. abnorm. soc. Psychol.*, 1956, **53**, 328–333.
6. EYSENCK, H. J. Cortical inhibition, figural after-effect, and theory of personality. *J. abnorm. soc. Psychol.*, 1955, **51**, 94–106.
7. FRANKS, C. M. Conditioning and personality: A study of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1956, **2**, 143–149.
8. GUILFORD, J. P. *An inventory of factors STDCR.* Beverly Hills, Calif.: Sheridan Supply, 1942.
9. HAMMER, ELAINE R. Temporal factors in figural after-effects. *Amer. J. Psychol.*, 1949, **62**, 337–354.
10. HOYT, C. J. Test reliability estimated by analysis of variance. *Psychometrika*, 1941, **6**, 153–160.
11. HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
12. KIMBLE, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
13. KIMBLE, G. A. Performance and reminiscence in motor learning as a function of the degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500–510.
14. KIMBLE, G. A., & BILODEAU, G. A. Work and rest as variables in cyclical motor learning. *J. exp. Psychol.*, 1949, **39**, 150–157.
15. KIMBLE, G. A., & HORENSTEIN, B. R. Reminiscence in motor learning as a function of length of interpolated rest. *J. exp. Psychol.*, 1948, **38**, 239–244.
16. KIMBLE, G. A., & SHATEL, R. B. The relationship between two kinds of inhibition and the amount of practice. *J. exp. Psychol.*, 1952, **44**, 355–359.
17. KÖHLER, W., & FISHBACK, J. The destruction of the Müller-Lyer illusion in repeated trials: II. Satiation patterns and memory traces. *J. exp. Psychol.*, 1950, **40**, 398–410.
18. KÖHLER, W., & WALLACH, H. Figural after-effects: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.
19. PAVLOV, I. P. *Lectures on conditioned reflexes*. Vol. II. London: Lawrence & Wishart, 1941.
20. SCHUCKER, R. E., STEVENS, L. B., & ELLIS, D. S. A retest for conditioned inhibition in the alphabet-printing task. *J. exp. Psychol.*, 1953, **46**, 97–102.
21. WERTHEIMER, M. Figural after-effects in schizophrenic and normal subjects. *Amer. Psychologist*, 1954, **9**, 492. (Abstract)
22. WERTHEIMER, M. The differential satiability of schizophrenic and normal subjects: A test of a deduction from the theory of figural after-effects. *J. gen. Psychol.*, 1954, **51**, 291–299.
23. WERTHEIMER, M. Constant errors in the measurement of figural after-effects. *Amer. J. Psychol.*, 1954, **67**, 543–550.

Received July 31, 1957.

BODY IMAGE AND ASYMMETRY OF BODY REACTIVITY[1]

SEYMOUR FISHER

Baylor University College of Medicine

In studies (6, 7, 8) focused on the manner in which the individual experiences his body boundaries, it was shown that persons differ greatly in the degree to which they conceive of their body boundaries as definite and firm as contrasted to indefinite and weak. One of the most meaningful of the observed correlates of degree of boundary definiteness had to do with site of psychosomatic symptomatology. The data indicated that persons who develop psychosomatic symptoms in the exterior body layers (e.g., neurodermatitis) are characterized by definite body-image boundaries, whereas individuals with body interior symptoms (e.g., stomach ulcers or spastic colitis) are marked by poorly defined body-image boundaries. Further, degree of boundary definiteness appears to play a determining part in this relationship and is not simply a reflection of the experiencing of pain and discomfort at particular body sites.

When it was found that body-image boundary definiteness was linked with body symptomatology, it followed that it may also be linked with more basic physiological variables. If one assumes that the development of psychosomatic symptoms at a given site indicates generally an exaggerated physiological reactivity or sensitivity at that site, one may hypothesize that prior to the development of obvious clinical pathology, individuals with definite boundaries are likely to manifest a higher degree of reactivity in their exterior body layers (skin, muscle, and peripheral blood vessels) and a lower degree of body interior reactivity than persons with indefinite boundaries. It was predicted that people with definite boundaries typically yield a gradient of physiological reactivity such that the body exterior is relatively more responsive than the body interior. Persons with indefinite boundaries were hypothesized to show a reversal of this gradient such that greatest reactivity to stimuli occurs in the body interior and relatively less in the exterior body layers. Substantiation for this hypothesis was obtained in two studies (7, 8) relating measures of boundary definiteness to measures of physiological responses at different body levels (e.g., MAP, GSR, heart rate).

This relationship of exterior-interior gradients of reactivity and an index of body image led to the proposition that one's body may be conceptualized as a screen upon which body-image attitudes are projected and that the pattern of this projection may influence the pattern of physiological reactivity. Thus, in applying this proposition to the body-image boundary dimension, one may hypothesize that the more definite and firm an individual's concept of his body boundary, the more likely he is to build up organized tonus sets and systems involving readiness to respond in the exterior body layers. In an oversimplified way, this process may be compared to the high tonus level that develops in an arm muscle when one arrives at a definite conclusion about a course of action that involves an arm movement and is about to initiate that movement. One's long term attitudes toward various sectors of one's body may result in selective modes of response in those sectors which significantly affect their physiological characteristics.

If this model concerning the relationship of body image to body reactivity is generally sound, it should be applicable to other than the boundary dimension of the body image. The principal purpose of the present study was to examine the generality of the body projection model by considering a new set of variables.

The new body image dimension studied had to do with the differentiation between the left and right sides of the body. There are indications in the literature on laterality (9, 10, 13, 18), stuttering (3, 12), and body delusions associated with schizophrenia (5, 14), and organic brain pathology (4, 14) that each individual, as he learns to differentiate between the right and left sides of his body, assigns definite values and significances to each side. The kinds of values which are customarily as-

---

[1] This project was supported by USPH Career Research Investigator Grant M-1578.

signed have not been ascertained in any systematic objective fashion. Some have speculated (5, 21) that one side is typically perceived as strong and the other as weak. Association may also be formed between sides of the body and masculinity-femininity, goodness-badness, and the acceptable aspects of the self vs. the unacceptable aspects of the self, etc. In any case, there is some basis for regarding the right-left differentiation as a potentially significant body-image dimension. According to the body projection model, it was hypothesized that the more clearly an individual distinguishes his right and left sides in his body image, the greater is the probability that he differs in physiological reactivity on the two sides of his body.

METHOD

Measure of the Right-Left Body-Image Distinction

To test this hypothesis, it was first necessary to devise a method for measuring how clearly an individual distinguishes between his two body sides. A continuum of bigness vs. smallness was selected because it seemed to embrace roughly a number of the left-right polarities about which there has been speculation in the literature (e.g., strong vs. weak, masculine vs. feminine). The specific method that was worked out for the measurement of this right-left difference involved having each *S* make a series of comparisons of the sizes of homologous fingers of the two hands. The comparisons are made while *S* wears a set of aniseikonic lenses of 4.00% power. In a group of ten cases, there were no consistent differences in actual size between the homologous fingers of the left and right hands. Furthermore, in those instances where some differences were observed, they were of the order of one or two millimeters and could not be reliably detected without a measuring scale.

The use of aniseikonic lenses was stimulated by the work of Wittreich[2] (20), who discovered that when an *S* wears the lenses and views himself in a mirror, it creates an unstable and moderately unstructured perceptual field into which it is relatively easy for one to project unconscious feelings about his body. One of Wittreich's over-all conclusions from his data was that individuals most resist the perception of novelty or distortion in those body areas about which they have the greatest anxiety or conflict.

In the present experiment, *S* was asked to don the lenses, which were set in spectacle frames, and told that they were designed to produce various distortions in the way things looked. Questions were asked concerning the spontaneous perceptual peculiarities experienced while looking around the room. When it was apparent that *S* was beginning to experience definite distortions, a series of specific perceptual alterations were demonstrated to him. These demonstrations involved the experiencing of tilt in the walls of the room, the perception of alteration in the shape of a rectangular bookcase, and the discovery of reversal in the slope of a pointer which seemed to be pointing toward the individual even though it was actually pointing away from him at a 45° angle. *S* was then told that the experimenter was interested in the kinds of distortions the lenses would produce while he was comparing the sizes of the fingers on his two hands. He was instructed to compare in turn the five sets of fingers, beginning with the two thumbs and ending with the two little fingers. He was shown how to pair each set of fingers and to make a judgment of which was larger when viewed at arm's length. It was emphasized that the differences involved were very small and that even the most minute size discrepancy between any two fingers should be reported. The intent was to indicate that *S*'s judgment was not reality bound and that he should try to experience a difference whenever possible. It was expected that the person with a well crystallized body-image distinction between the right and left body sides would tend to perceive differences in finger sizes and that the differences would be consistently in a given direction. On the other hand, the person with poorly articulated right-left body-image distinctions was expected to be relatively conflicted and anxious about the size judgments and therefore either unable to experience differences at all or able to experience only inconsistent differences. *S*'s score for the aniseikonic judgments could vary at one extreme from seeing no differences between any of the sets of fingers to the other extreme of seeing each of the five fingers on a given body side as larger than the corresponding fingers on the opposite side.

Measure of Right-Left Body Reactivity

The GSR was selected as the index of reactivity in the two body sides. It was chosen primarily because it seemed to lend itself better than other measures to carefully controlled bilateral recordings. In order to compare the physiological reactivity of the right and left sides of the body, it was necessary to use a technique which would permit recordings that could be compared with full confidence that any differences would not be a function of differences in the apparatus associated with the measurement process on either side. The method of GSR recording was of the exosomatic type. With a constant current of 20 microamperes, resistance changes across the skin were picked up as changing potential and fed into a Grass EEG power amplifier.[3] This type of recording produces a write-out that is more a derivative of the GSR than a true wave form. But calibration against DC recording showed a linear relationship between the two. Recordings from the right and left side were made simultaneously on

---

[2] Wittreich, W. J. Body image and psychopathology: A pilot study. Unpublished report to Office of Naval Research, 1954. Wittreich, W. J. and Grace, M. Body image development. Unpublished progress report to Office of Naval Research, 1954.

[3] It is acknowledged with gratitude that Robert Edelberg devised and constructed the entire GSR recording system.

TABLE 1

DISTRIBUTION OF SIZE JUDGMENTS IN WHICH A MAJORITY OF FINGERS ON ONE HAND JUDGED AS LARGER THAN FINGERS ON OTHER HAND

	Right-handed			Left-handed		
	Male	Female	Total	Male	Female	Total
Majority of rights larger[a]	19	13	32 (59%)	2	6	8 (33%)
Majority of lefts larger	5	0	5 (9%)	3	5	8 (33%)
No majority	11	6	17 (31%)	6	2	8 (33%)

[a] Three or more fingers on a given hand judged as larger.

TABLE 2

DISTRIBUTION OF SUBJECTS ON SIZE JUDGMENTS CONCERNING RIGHT- AND LEFT-HAND FINGERS

	Right Fingers Larger		Left Fingers Larger	
	Right-handed	Left-handed	Right-handed	Left-handed
Number judging above median[a]	25	3	22	18
Number judging at median or below	29	21	32	6

[a] The median for the combined right and left-handed groups.

separate channels against a relatively inactive upper arm. A very large lead electrode was used on the upper arm to minimize any response in that area. Recordings were taken simultaneously first from the middle fingers; second, from the forefingers; third, from the ring fingers, and, finally, from two corresponding points on the palms. The currents to each finger were carefully matched. Also, a common calibrating resistor in series with the upper arm checked the equivalence of the sensitivities of the two circuits. The area of recording from each site was made equivalent by using pieces of tape with holes of ¼-in. diameter punched into them. The piece of tape was fixed on the recording site; a small quantity of electrode jelly was applied to the skin area enclosed by the hole; and a lead electrode was taped into place over the hole. At least two minutes of recording were obtained from each set of sites. Longer periods were monitored if it was difficult to elicit GSRs from particular *S*s. If *S* did not spontaneously produce responses, he was aroused by making loud noises, asking him to grit his teeth, suddenly touching him, etc. For each *S*, there was finally available a continuous side-by-side record of GSRs for each set of right-left sites. The amplitude of response occurring at two homologous sites at a given time could then be determined and compared. So far as can be ascertained, such simultaneous balanced recording of bilateral GSR responsivity for the purpose of studying body reactivity gradients has not been previously reported.

Subjects

There were 78 *S*s, 46 male and 32 female. Ages ranged from 22 through 38, and the median educational level was 14 years. Twenty were acquaintances of the experimenter who agreed to participate, and 58 were college students who were asked to volunteer and offered five dollars for their services.

From the very outset of the experiment, it became apparent that hand dominance was an important variable. Hand dominance was evaluated in all cases. A very approximate method was used of simply inquiring which hand the *S* used for various important activities (e.g., writing, eating, athletic skill, tasks requiring unusual strength). Fifty-four *S*s (35 male, 19 female) were right-handed and 24 (11 male, 13 female) left-handed.

RESULTS

When the judgments concerning finger size were tabulated, as shown in Table 1, it was found that 59% of right-handed *S*s considered a majority of their right-hand fingers larger than their left-hand fingers. Nine per cent considered a majority of their lefts larger, and 31% gave a majority to neither right nor left. Although the tendency was to see a majority of right-hand fingers as larger than left, 40% of *S*s either reported no consistent differences between the right- and left-hand fingers or perceived the left-hand fingers as relatively larger. The left-handed *S*s, as shown in Table 1, distributed their judgments of relative finger size equally into the categories right, left, and no difference. They did not seem to attribute obvious superiority in size to the dominant hand in the same degree as the right-handed *S*s.

A chi-square comparison of the number of *S*s in each hand dominance group who judged a majority of right-hand or left-hand fingers as larger yielded a value of 5.9, significant between the .02 and .01 levels. None of the differences among sex groups were significant.

In Table 2 is presented another analysis of the distribution of judgments concerning size of right- vs. left-hand fingers in the hand dominance groups. These data are concerned with the median number of fingers judged larger or smaller on each hand. A considerably greater proportion of right-handers than left-handers perceived an above-median number of right-hand fingers as larger than left-hand fingers. The difference, as evaluated by chi square, is significant at the .01 level. Although the left-handed *S*s did not seem to attribute clear-cut superiority in size to a majority of left-hand fingers, Table 2 indicates that they exceeded right-handed *S*s in the proportion of left-hand fingers described as larger. This difference, as determined by chi square, is significant at the .01 level.

In evaluating the GSR data, the criterion of directionality of response was based on a comparison of the left and right amplitudes given at each site. All corresponding responses in which a difference of at least 80 ohms occurred were noted. This level of difference was chosen because it represented 2 mm. on the measuring scale and was the smallest that could be reliably measured with a compass. Even in the maximum possible response, 2 mm. still represented 8% of the measured range. If the differences occurring for a given site were such that there were twice as many favoring one side as the other, and if at least ten differences were present, then it was considered that a directionality of response for that site had been established. Further, when *S* manifested this sort of directionality in favor of one side for three out of the four sites, he was designated as having a definite GSR gradient in the given direction. If less than three of the four sites were consistently in one direction, it was judged that a definite right-left difference was not present. For 43 of the 54 *S*s who showed a definite directionality of response, the ratio of predominant differences to nonpredominant differences at each site was 3 to 1 or greater. Also, for 26 of these 54 *S*s, all four of the sites showed predominant responsivity in a given direction. The reliability of the directionality of differences was explored by retesting three right-handers after intervals ranging from two weeks to two months. In all three instances, the retest directionality of response was exactly the same as that obtained for the first testing.

As can be seen in Table 3, individuals tend to show definite gradients of GSR reactivity, such that either one side responds to a relatively greater degree than the other. Only 24 of the 78 *S*s failed to evidence a clear-cut gradient. This finding is in line with Von Helmut Baitsch's (1) report that in serial exploration of the skin resistance of the fingers on the two hands, 60 to 80% of his *S*s had higher resistance on the right side than on the left.

There is a tendency suggested in Table 3 for right-handed women to exceed right-handed men in the relative frequency with which they manifest a definite left directionality in their GSR responses. But the chi-square value, when corrected for discontinuity, falls only at the .10–.05 level.

TABLE 3

DISTRIBUTION OF GSR RESPONSIVITY DIRECTIONALITY IN RIGHT-HANDED AND LEFT-HANDED SUBJECTS

	Right-handed			Left-handed		
	Male	Female	Total	Male	Female	Total
Right responsive	7	5	12	5	2	7
Left responsive	14	12	26	2	7	9
No clear gradient	14	2	16	4	4	8

In order to check the basic hypothesis that difference in physiological reactivity between the two body sides is a function of how definitely the two sides are differentiated in the body image, it was necessary to relate the aniseikonic finger size judgments to the GSR indices of response directionality. This was done for each of the hand dominance groups separately because it was apparent by inspection that a different pattern of relationship occurred in each. (Incidentally, the relationship between GSR directionality and lateral dominance was not in itself significant.) Within the right-handed group, the median number of right-hand fingers judged as larger than left-hand fingers was 3. The median number of left-hand fingers judged as larger was 0. Further, the median number of right-hand fingers judged larger was 2 in the group showing no distinct GSR right-left differences, 2 in the group characterized by right GSR directionality, and 4 in the group typified by left GSR directionality. *S*s in the total right-handed group were divided into those above the median in number of right-hand fingers judged as larger and those below the median in this respect; this categorization was related by means of chi square to left directionality of GSR response vs. either no directionality or right directionality. Those above the median in their aniseikonic differentiation of their right and left sides showed a greater tendency toward left GSR directionality than did those below the median. Chi square was significant beyond the .001 level.

A word is in order concerning the manner in which GSR directionality was categorized. When a GSR gradient was present, it usually involved relatively greater reactivity on the side judged as small than on the side described as large. Categories were established accordingly which treated left reactivity as the norm and other reactivity patterns as somehow deviant.

In the left-handed group, the median number of left-hand fingers judged larger than right-hand fingers was 2. The median number of right-hand fingers judged larger than left-hand fingers was 1. The median number of left-hand fingers judged as larger was exactly the same in the GSR right reactive, left reactive, and no directionality categories. *S*s in the left-handed group were divided into those above the median in number of left-hand fingers judged as larger and those below the median in this respect. Chi square was then utilized to determine whether these groups differed with respect to the right directional GSR category vs. the combined category of either no directionality or left directionality. The results were not significant. The GSR categories which were employed duplicated the logic of those used in the analysis of the right-handed group. When the aniseikonic judgments in the left-handed group were related to other modes of categorizing GSR reactivity, only change relationships were found.

The over-all results indicate, then, that in the right-handed group those who make the clearest aniseikonic distinction between right and left sides are those most likely to show GSR reactivity on the left side. But right-handed *S*s who distinguish poorly between their right and left sides are likely to be right reactive or to have no definite gradient at all. Among the left-handers, there was no evidence of a relationship between the aniseikonic lens judgments and directionality of GSR response.

DISCUSSION

How has the original hypothesis fared? Originally, there was no consideration of the possibility that laterality might complicate the relationships under study. It would appear, however, that only when laterality is taken into account is there evidence for the relationship between body image and body reactivity which was asserted. One may conjecture from the pattern of results that the body-image variable defined by the aniseikonic lens technique exerts its influence upon bodily reactivity within the setting of gradient effects previously established in the process of developing lateral dominance. In situations uncomplicated by body-image variables, right-hand dominance may be associated with left directionality of GSR response and left-hand dominance with the opposite GSR directionality. The experiences which determine the clarity of differentiation in the body image of the right and left body sides may occur after the basic laterality gradient has been built up, producing an impact on right vs. left reactivity within the context of the pre-existing laterality gradient. For the average person, the basic process of right-left body-image differentiation possibly involves the assigning of different values to the two body sides in such a way that attributes of bigness are projected on to the dominant side and attributes of smallness on to the other. If for some reason the division of attributes cannot be made clearly or if bigness is assigned to the nondominant rather than to the dominant side, there may be a disruption or even a reversal of the existing reactivity gradient based on laterality. At this point one can only speculate about the process involved.

The more immediate question is that of why the body-image variable is significantly related to directionality of GSR reactivity in the right-handed group but not in the left-handed group. In a sense, this question, which concerns a contrast between right and left-handed persons, is a familiar one in the literature concerned with laterality. Several studies have found right-handed persons to be more homogeneous and consistent with regard to the types of variables under consideration than left-handed persons (2, 10). An excellent and particularly pertinent study in this respect is one reported by Wishner and Shipley (19). They were interested in the effects of laterality upon the right-left directionality of autokinetic movement. They found that right-handed *S*s exhibit a consistency tendency to perceive movement toward the right, whereas left-handed *S*s evidence no consistency in either direction. It seems that left dominance is almost always far less definite than right dominance and that tension patterns associated with laterality are more irregular in the left-handed than in the right-handed. On this basis, one may speculate that the relationship between right-left body image differentiation and right-left reactivity is relatively attenuated in the left-handed group because of the relatively diffuse and complicated character of left dominance.

Another important question has to do with the tendency for GSR reactivity to be greater on the body side perceived as the smaller. The model for explanation that comes to mind is based on the idea that the side experienced as larger (and which in right-handers with a definite GSR gradient is usually the dominant side) is assigned the function of guiding and controlling motor responses to problem situations. The executive monitoring function of one hand in an on-going task is, of course, a well confirmed phenomenon (10). The integrity and stability of this set for directive action may be of considerable psychological importance. Therefore, mechanisms may be developed for preventing surges of excitation which could disrupt or significantly alter the set. This possibility suggests the hypothesis that disruptive excitation is "drained off" or channeled into the nondominant side, where its effects are relatively less disorganizing. Thus, the nondominant side is likely to show relatively greater response to many kinds of excitation than is the dominant or "large" side. This formulation implies an optimum homeostatic[4] relationship between the two body sides. The tendency for GSRs to occur on the "small" side thus reflects the relatively greater degree of extraneous excitation diverted there in order to maintain the optimum set on the other side.

Previous work by Schoen and Scofield (15), Travis and Herren (17), and Jasper (11) lends support to this model. Schoen and Scofield found that when the eyes of a right-handed person were shifting from one fixation point to a new target, the left eye responded first, "snapping" to its new position and sometimes overshooting, as contrasted to the right eye, which proceeded more gradually and in a more coordinated fashion. Travis and Herren and Jasper observed that in situations where right-handed individuals were instructed to perform some simple task quickly and simultaneously with both arms, the left arm tended to respond first. These three studies point up the reality of a right-left difference in motoric reactivity and favor the view that the dominant side is characterized by a stable set which facilitates control but which also inhibits the spontaneity of that side as compared to the nondominant side. Jasper (11, p. 161) proposes that right-left differences in cortical organization may be responsible:

> The dominance of the cortex in the control of reflex movement is a commonly accepted principle for the explanation of the more highly integrated activity of the central nervous system. If this principle is applied to lateral dominance the tendency for the non-dominant side to lead in attempted simultaneous movement may indicate a greater cortical control ("inhibition") of the movements of the so-called dominant side which is only a counterpart of the more highly perfected coordination of movement on this side.

But the relationship found here between the right-left body image index and right-left GSR reactivity indicates that the differential in lateral response referred to by Jasper is not simply a function of lateral dominance factors. It may also be influenced by the manner in which the individual has evolved a right-left division in his body image.

The general findings essentially support the hypothesis tested. It is, of course, true that the hypothesis was oversimplified and neglected the laterality factor. However, what is outstanding is that within the right-handed group, a highly significant linkage (.001 level) was observed between the individual's body-image pattern of right-left differentiation and his pattern of right vs. left physiological reactivity. The significance of these findings increases when they are considered in the context of previously reported work (8, 9) which demonstrated that the patterning of physiological reactivity in the exterior vs. the interior of the body may be predicted from the manner in which the individual conceives of his body boundaries. At two different levels, then, body-image variables seem predictive of physiological gradient variables. The possibility is impressive that body-image concepts may prove to be one of the best means for bridging the gap between psychological and psychodynamic formulations concerning the individual and his patterns of physiological reactivity. The concepts and techniques described in the present paper lay the groundwork for investigating with considerable precision the nature of various "psychosomatic" relationships.

---

[4] Acknowledgement is made to Neil Burch for his helpful suggestions concerning the issue of lateral homeostasis.

SUMMARY

The hypothesis was proposed that the more definitely an individual distinguishes the right and left sides of his body in his body image, the more likely he is to manifest differential reactivity in the left vs. right sides of his body. Body-image differentiation of right and left sides was studied by means of a technique requiring *S*s to compare homologous body parts while viewing them through aniseikonic lenses. Right vs. left body reactivity was measured by simultaneous and balanced GSR recordings from the two body sides. It was found that among the right-handed *S*s those clearly distinguishing their right and left body sides were likely to show a GSR gradient such that the left side was more reactive than the right. Those without a clear body-image distinction between the right and left body sides either showed no GSR gradient or were predominantly right reactive. Among the left-handed *S*s there was no observable relationship between degree of right-left body-image differentiation and right vs. left GSR. It is considered that the data basically support the hypothesis tested.

REFERENCES

1. Baitsch, H. Über Geschlechts- und Seitendifferenzen im "Niveau"-Elektrodermatogramm. *Conf. neurol.*, 1954, **14,** 88–100.
2. Bauer, R. W., & Wepman, J. M. Lateralization of cerebral functions. *J. Speech Dis.*, 1955, **20,** 171–177.
3. Bolin, B. J. Left-handedness and stuttering as signs diagnostic of epileptics. *J. ment. Sci.*, 1953, **99,** 483–488.
4. Critchley, M. *The parietal lobes.* London: Edward Arnold, 1953.
5. Fenichel, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.
6. Fisher, S., & Cleveland, S. E. The role of body image in psychosomatic symptom choice. *Psychol. Monogr.*, 1955, **69,** No. 17 (Whole No. 402).
7. Fisher, S., & Cleveland, S. E. An approach to physiological reactivity in terms of a body-image schema. *Psychol. Rev.*, 1957, **64,** 26–37.
8. Fisher, S., & Cleveland, S. E. *Body image and personality.* Princeton, New Jersey: Van Nostrand, in press.
9. Hewes, G. W. Lateral dominance, culture, and writing systems. *Hum. Biol.*, 1949, **21,** 233–245.
10. Hildreth, G. The development and training of hand dominance: IV. Developmental problems associated with handedness. *J. genet. Psychol.*, 1950, **76,** 39–100.
11. Jasper, H. H., & Raney, E. T. The physiology of lateral cerebral dominance. *Psychol. Bull.*, 1937, **34,** 151–165.
12. Lindsley, D. B. Bilateral differences in brain potentials from the two cerebral hemispheres in relation to laterality and stuttering. *J. exp. Psychol.*, 1940, **26,** 211–225.
13. Martin, K. L. Handedness: A review of the literature on the history, development, and research of laterality preference. *J. educ. Psychol.*, 1952, **45,** 527–533.
14. Schilder, P. *The image and appearance of the human body.* New York: International Univer. Press, 1950.
15. Schoen, Z. J., & Scofield, C. F. A study of the relative neuromuscular efficiency of the dominant and non-dominant eye in binocular vision. *J. gen. Psychol.*, 1935, **12,** 56–181.
16. Thass-Thienemann, T. Left-handed writing. *Psychoanal. Rev.*, 1955, **42,** 239–61.
17. Travis, L. E., & Herren, R. Y. Studies in stuttering. V. A study of simultaneous antitropic movements of the hands of stutterers. *Arch. Neurol. Psychiat.*, 1929, **22,** 487–494.
18. Wile, I. S. The relation of left-handedness to behavior disorders. *Amer. J. Orthopsychiat.*, 1932, **2,** 44–57.
19. Wishner, J., & Shipley, T., Jr. Direction of autokinetic movement as a test of the "Sensory-Tonic-Field" Theory of perception. *J. Pers.*, 1954, **23,** 99–106.
20. Wittreich, W. J., & Radcliffe, K. B., Jr. The influence of simulated mutilation upon the perception of the human figure. *J. abnorm. soc. Psychol.*, 1955, **51,** 493–495.
21. Wolff, W. *The expression of personality.* New York: Harper, 1943.

Received August 3, 1957.

TIME ORIENTATION AND INTERPERSONAL RELATIONS OF EMOTIONALLY DISTURBED AND NORMAL CHILDREN[1]

ANTHONY DAVIDS AND ANITA NEGRIN PARENTI

Brown University and Emma Pendleton Bradley Hospital

THE aim of this research was to investigate time orientation and interpersonal relations in a group of hospitalized emotionally disturbed children and in two groups of normal children. Time orientation in imaginal processes was studied by a projective method, and interpersonal relations were defined by measures of stability of friendship patterns. These two significant areas were previously investigated independently (1, 5, 7). The present study was designed to replicate and extend these previous studies and to relate and unify them.

In a study of time orientation and social class, LeShan (6) found that lower-class children tend to be present-oriented, while middle-class children tend to be future-oriented. In a related study, Barndt and Johnson (1) compared time orientation in groups of delinquent and nondelinquent boys who were equated for socioeconomic status. They found the delinquents to be significantly more present-oriented than were the nondelinquents. In both of these investigations, it was speculated that emotionally disturbed individuals have short-time perspectives. On the basis of a recently reported experiment with psychotic adults, Wallace stated that "... future time perspective is influenced by the schizophrenic process to such an extent that both the length of the future time span and the organization of its contents are significantly reduced for a sample of schizophrenic patients as compared to a group of normal controls. Additional research is needed, however, before such a finding can be considered to have been conclusively demonstrated" (8, p. 245).

Two studies on stability of friendship patterns, carried out by Thompson and Horrocks (5, 7), are also directly relevant. While other inquiries focused on time orientation as assessed by responses to a projective technique, Thompson and Horrocks investigated fluctuations in friendship choices in everyday social relations. With both rural and urban children, they found that in 11-year-olds there was considerable fluctuation in friendship choices over a two-week period of time. However, with increasing age there was an increasing degree of stability in friendship patterns. Thus, it would appear that stability of friendship patterns is a correlate of emotional stability.

Goldfarb (4) showed that children who are emotionally deprived in their early years and who are reared in institutional settings tend to have deficient time orientation. He argued that this deficiency is definitely a factor in their limited foresight and their inability to envision the consequences of their acts and is even a major cause of their social maladjustment.

It appears, then, that time orientation is crucial both to emotional and social adjustment. The present study examines this notion empirically.

METHOD

Subjects

The experimental *S*s in this study consisted of a group of 30 hospitalized emotionally disturbed boys undergoing residential treatment at the Emma Pendleton Bradley Hospital in Riverside, Rhode Island. They ranged in age from 7 years, 5 months, to 13 years, 4 months, with a mean age of 10 years, 10 months. The children in this group were suffering from severe behavior disorders, neuroses, or psychoses. Most of them had been at the Bradley Hospital for a period of not less than eight months, and the average length of hospitalization was approximately two years. The boys represented a heterogeneous sample with respect to socioeconomic status, but most of them were from middle- and lower-class families.

---

[1] We wish to express our appreciation to the director and the child care staff of the Bradley Hospital for their cooperation in our work with the disturbed children. We are also indebted to Edward E. Martin, superintendent, and Louise Favorite, psychologist, for their cooperation in our study of children in the East Providence, Rhode Island public schools. And Murray Ortof, director, was very helpful in our study of children at Camp Hurley in Kingston, New York. We are also grateful to Sylvia Topal, staff social worker at the camp, and Johanna Walter, psychiatrist at Bradley, who provided clinical ratings of the children. This research was facilitated by a summer undergraduate research stipend awarded to Anita Negrin Parenti from the Social Science Research Council, and by Research Grants M-1722 and M-2087 awarded to Anthony Davids from the National Institute of Mental Health, Public Health Service.

One group of 33 control *S*s was drawn from a population of children attending a summer camp for a period of three weeks. The camp (Camp Hurley, in Kingston, New York) took in children from all socioeconomic and ethnic groups. Although social workers were employed as supervisors to the counselors, who lived and worked in direct contact with the children, the camp was for normal youngsters. The boys studied ranged from 7 years, 11 months, to 13 years of age, with a mean of 10 years, 10 months.

Another control group of 22 boys was obtained from the Grove Avenue School in East Providence, Rhode Island. The children were in the fourth, fifth, and sixth grades. They ranged from 9 years, 8 months, to 12 years, 4 months of age, with a mean of 10 years, 10 months.

Thus the *S*s were comparable for age, and although they were not, strictly speaking, equated for socioeconomic status, there is no reason to believe that any systematic class differences existed among the groups.

Time Orientation in Fantasy

The story-completion technique. The procedure followed was previously employed and described by Barndt and Johnson (1) and by Wallace (8). The *S* was given the following instructions: "I want to see what kind of a story you can tell. I'll start a story and then let you finish it any way you want to. You can make it any kind of a story you wish. Do you understand? I'll start it now." All four of the following stories were administered to the Bradley sample, the first two to the camp sample, and only the first one to the school sample. The first story began as follows: "About three o'clock one bright sunny afternoon in May, two boys were walking along a street near the edge of town. . . ." The second story began: "Ten o'clock one morning Al met his friend Jerry near the center of town. . . ." The beginning of the third story was: "Two boys were walking along the street three o'clock one sunny afternoon. One boy said to the other. . . ." And the fourth story began: "Johnny and Jimmy met each other ten o'clock one morning. . . ." After reading the beginning of each story to the child, *E* told him, "Now you start there and finish the story any way you want to." *E* then recorded the child's story verbatim. After finishing the story, the child was asked, "What time was this?" or "How long did the story take to happen?" In the cases where children were unable to report the length of time covered by their stories, *E* estimated the time span, but this occurred very infrequently.

In the Bradley sample, the first and second stories were administered during the first interview and the remaining two after two weeks. In the camp sample, only the first story was given during the first interview; the second was given three weeks later.

Scoring of stories. The scoring system described by Barndt and Johnson (1) was used. The scoring categories were as follows:

1. Under one hour
2. One hour to under five hours
3. Five hours to under twelve hours
4. Twelve hours to under one week
5. One week to under three months
6. Three months or more

The amount of time covered in the stories could be scored very accurately, and agreement between raters was unity.

Reliability of measures. Using the six-point scale described above, a product-moment correlation of .75 between the first and second stories was obtained for the Bradley sample. Correlations of .66 and .73 were found between the second and third and the third and fourth stories. In the camp sample, a correlation of .93 was obtained between the first and second stories. These coefficients indicate adequate consistency in these measures of time orientation. Only scores from the first story were used in the analyses reported here.

Stability of Interpersonal Relations

The friendship fluctuation method. In assessing friendship patterns, the method used by Thompson and Horrocks (5, 7) was adapted. The child was told: "Please tell me the names of your three best friends, or the children you like best at Bradley (camp or school). Tell me the one you like best, then the second best, and then the third best." The children were asked to name their friends on four different occasions, with a one-week interval between each interview. At each of the subsequent interviews, the original instructions were repeated and the following instructions were added: "The people you tell me may or may not be the ones you liked best the last time. Be sure to tell me the three you like now."

Scoring of friendships. To score the friendship patterns, a point system was taken from Thompson and Horrocks (5, 7) with a few modifications. One point was given for each change in position of a friend or friends. For example, if John were selected during the first interview as the very best friend of *S* and if during the next interview he was selected as *S*'s third best friend, two points were awarded. If John were selected as *S*'s second best friend during the second interview, one point was awarded. Seven points were awarded for the selection of a best friend who did not appear in the choice pattern during the previous interview. Similarly, six points were awarded for the selection of a second best friend and five for the selection of a third best friend who did not appear in the previous choice pattern.

The highest score that could be obtained by retaining the same friends but switching the order of preference was four. It was felt that a higher degree of instability was indicated by choosing an entirely new friend or friends than was indicated by switching the order of preference of previously named friends. More points were therefore given for the appearance of a new friend than for the maximum changing of preferences.

By means of this scoring system, it was possible to obtain several friendship fluctuation scores for each child: (*a*) over three weekly periods, (1–2, 2–3, and 3–4); (*b*) over two two-week periods, (1–3, and 2–4); and (*c*) over one three-week period, (1–4). Except where otherwise noted, however, each child's friendship fluctuation score represents the mean of the three weekly periods. This score was used in computing all correlations. For purposes of comparison, the friendships were rescored according to the same method used by Thompson and Horrocks (5, 7). This procedure entailed scoring the change in friendships over a two-week period. Four points were awarded for the appearance of a new best friend in the choice pattern, three points for the appearance of a new second best friend, and two points for the appearance of a new third best friend. In addition, one point was awarded for each change in position of a friend or friends. In the statistical treatment of the data, it was felt that it was better to consider the friendship fluctuations of weeks two-to-four in the camp group as being comparable to weeks one-to-three in the Bradley and school groups. This was done because the children in the camp group had not known each other for more than one day when the first friendship measure was taken, while those in the Bradley and school groups had known one another for some time prior to the study. However, this method was used only in comparing the camp group to the children studied by Thompson and Horrocks (5, 7).

Measurement of Emotional Disturbance

The rating scales. Rating scales were constructed for each of the following eight personality dispositions: sociocentricity, optimism, trust, egocentricity, pessimism, distrust, anxiety, and resentment. Of these traits, the first three are positive or socially desirable; the last five are negative or socially undesirable (2). In addition to ratings on the individual traits, a scale was constructed on which *S* could be rated in terms of the degree of his "general emotional maladjustment." Detailed definitions of the variables were included on the mimeographed sheets containing the rating scales. The raters were instructed to judge each *S* in comparison with other children of his age and sex in his social group. The ratings were made on six-point scales running from "much less than most" to "much more than most," with a score of one being assigned to the lower extreme and a score of six assigned to the higher extreme. Thus, it was possible to utilize scores on the individual dispositions, sums on the three positive or on the five negative dispositions, or scores on the general rating.

Reliability of ratings. Ratings were obtained for the Bradley sample and the camp sample. The children at Bradley were rated by a staff clinical psychologist and a staff psychiatrist. The children at camp were rated by counselors and a professional social worker. The product-moment correlation between the ratings made by the psychologist and psychiatrist on general emotional maladjustment for the Bradley sample was .77, while the comparable correlation between the counselors and social worker at camp was .54. In the statistical comparisons contained in this report, the ratings by the psychologist (Bradley) and by the social worker (camp) were employed.

HYPOTHESES AND RESULTS

The hypotheses to be tested in this experiment were derived from the theoretical discussions and reports of previous investigators who have found present orientation and instability of friendships to be correlates of immaturity and emotional maladjustment.

Hypothesis I: *Disturbed children are more present-oriented in fantasy than are normal children.* It is evident from Table 1 that this hypothesis was not confirmed. Actually, the disturbed group received a slightly higher, though nonsignificant, score on future orientation than did the two normal groups.

Hypothesis II: *Disturbed children show less stable friendship patterns than do normal children.* The findings in Table 1 indicate that the disturbed sample was characterized by a significantly higher friendship fluctuation score than either of the control groups, while the normal groups did not differ from one another. Thus, it appears that the patterns of interpersonal relations in the disturbed group are significantly more unstable than in the normal groups. The percentage of change in choice of best friend over a two-week period was also

TABLE 1

COMPARISON OF GROUPS ON STORY-COMPLETION AND FRIENDSHIP FLUCTUATION SCORES

Measure	Bradley ($N = 30$)		Camp ($N = 33$)		School ($N = 22$)		Bradley vs. Camp		Bradley vs. School		Camp vs. School	
	Mean	S^2	Mean	S^2	Mean	S^2	F	t	F	t	F	t
Friendships (Mean)	8.13	15.62	5.39	8.57	6.04	16.52	1.82	3.22**	1.06	1.86*	1.93*	.65
Friednships (1–2)	8.97	22.48	6.18	13.09	6.23	22.56	1.79	2.86**	1.04	2.04*	1.72	.04
Friendships (2–3)	8.00	20.90	5.48	17.45	6.91	33.13	1.20	2.36*	1.59	.76	1.90	1.07
Friendships (3–4)	7.43	29.22	4.52	12.57	5.00	18.57	2.32*	2.51**	1.57	1.74*	1.48	.45
Friendships (1–3)	9.07	21.93	7.00	18.63	7.45	24.45	1.18	1.88*	1.11	1.21	1.31	.36
Friendships (2–4)	9.80	31.48	5.79	20.48	6.73	27.06	1.54	3.23**	1.16	2.02*	1.32	.71
Friendships (1–4)	9.60	21.90	7.61	18.93	6.23	16.47	1.16	1.75*	1.33	2.70**	1.15	1.18
Friendships (1–3) Thompson Scoring	4.73	5.44	3.73	4.83	3.95	6.52	1.13	1.79*	1.20	1.15	1.35	.33
Story scores	2.93	2.69	2.39	1.31	2.86	1.36	2.05	1.50	1.98	.17	1.04	1.49

* Significant at the .05 level with a one-tailed test.
** Significant at the .01 level with a one-tailed test.

computed. The percentages obtained were 47, 42, and 59 for the Bradley, camp, and school samples respectively; there are no significant differences among these values.

Hypothesis III: *There is an association between time orientation in fantasy and stability in friendship patterns.* In the Bradley group, there is a significant positive correlation of .61 between friendship and story scores, while in the two normal groups there is no relation between the two measures. This indicates that with the emotionally disturbed *S*s, the more present-oriented the child is on the story-completion method, the more stable his friendship patterns tend to be.

Hypothesis IV: *Degree of general emotional maladjustment is correlated with time orientation and with stability of friendship patterns.* In the Bradley sample, there is no relation between these variables. In the camp sample, there is only a nonsignificant positive correlation of .22 between the ratings on general emotional maladjustment and story scores. In the normal sample, there is a significant positive correlation of .33 between the ratings on maladjustment and friendship fluctuation scores. This indicates that the more disturbed the child is, within the normal sample of children, the more likely he is to change his friends from week to week.

Hypothesis V: *Ratings on the positive personality traits are correlated with future orientation in fantasy and stability of friendship patterns, while ratings on the negative personality traits are correlated with present orientation and instability of friendships.* The ratings on the three positive traits (sociocentricity, optimism, trust) were summed to arrive at a general rating of good adjustment, as opposed to the previous rating on general emotional maladjustment used in testing Hypothesis IV. As can be seen from Table 2, Hypothesis V was not confirmed. For the Bradley sample, this positive sum bore no relation to either the story or friendship scores. In the normal sample, however, there is a significant correlation of −.41 between the sum of the positive traits and story scores, indicating that the more present-oriented a child is, the more likely he is to be rated high on positive traits. The correlation of −.22 between the sum of the positive traits and the friendship scores is nonsignificant.

Correlations between ratings on the individual personality traits and the story and friendship scores are shown in Table 2. The results indicate no significant relations between the story scores and the individual traits in the Bradley sample. In the camp sample, there is a significant correlation of .50 between pessimism and story scores, indicating that the children who were judged to be pessimistic tended to be future-oriented on the fantasy measure. The rating on resentment also correlated significantly (.41) with story scores, suggesting that the more resentful children were more future oriented And the significant correlation of −.43 between trust and story scores implies that the more trusting children tended to be more present-oriented in fantasy.

Of the correlations between personality traits and friendship fluctuations, the only significant coefficient (.48) in the Bradley sample suggests that children judged clinically to be high on resentment evidence greater fluctuations and instability in their choice of friends from week to week. In the camp sample, the significant correlation of −.30 is consistent with the notion that the more sociocentric a child is, the less fluctuating are his friendship patterns.

While these findings lack consistency, they

TABLE 2

CORRELATIONS BETWEEN MEASURES OF TIME ORIENTATION AND RATINGS ON PERSONALITY TRAITS

Group and Measure	Egocentricity	Pessimism	Distrust	Anxiety	Resentment	Sociocentricity	Optimism	Trust	Sum of Positive Ratings
Bradley (N = 30)									
Story scores	.02	−.19	.03	.10	.15	−.08	−.06	−.03	−.08
Friendship scores	−.06	−.10	.00	.03	.48**	−.07	−.11	−.08	−.10
Camp (N = 33)									
Story scores	.28	.50**	.33	.15	.41*	−.31	−.29	−.43*	−.41*
Friendship scores	.24	.22	.27	−.10	.25	−.30*	.02	−.25	−.22

* Significant at the 05 level.

** Significant at the .01 level.

are characterized by certain noteworthy trends. The ratings on the negative traits appear to be positively correlated with both story and friendship scores, while the ratings on the positive traits are, with only one exception, negatively correlated with these variables. This indicates that the more well-adjusted children tend to show less change in their friendships and tend to tell stories that cover a shorter time span.

In addition to tests of the hypotheses, some direct comparisons were made between the results obtained here and those reported by previous investigators. A comparison between the results of this study and Barndt and Johnson's (1) in their study of 17-year-old delinquent and nondelinquent boys was possible because the story beginnings and the scoring system were identical. The results presented in Table 3 show that all of our three groups of 11-year-olds were more present-oriented than their group of 17-year-old nondelinquents, whereas there were no differences between our groups and their delinquents.

The mean friendship fluctuation scores of the two normal groups in the present study, when computed for a two-week period by the exact method used by Thompson and Horrocks (7) in their study of urban children, did not differ significantly from the scores reported by these investigators. The degree of friendship fluctuation for the disturbed group, however, was greater than that reported for their normal 11-year-old boys. These investigators found approximately a 50% change, at this age level, in choice of best friends over a two-week period. As noted above, the change within each of the three groups in the present study was also approximately 50%.

TABLE 3

COMPARISON OF PRESENT THREE GROUPS WITH GROUPS STUDIED BY OTHER INVESTIGATORS

Group	M	S^2	Bradley		Camp		School	
			F	t	F	t	F	t
Barndt & Johnson Delinquents ($N = 26$)	2.8	1.96	1.37	.32	1.50	1.24	1.44	.16
Barndt & Johnson Controls ($N = 26$)	3.6	1.69	1.59	1.68*	1.29	3.78**	1.24	2.05*
Thompson & Horrocks 11-year-olds ($N = 21$)	3.48	4.62	1.18	2.03**	1.13	.38	1.41	.70

* Significant at .05 level.
** Significant at .01 level.

DISCUSSION AND CONCLUSIONS

While there were no significant differences between normal and disturbed boys with respect to time orientation as measured by the story-completion technique, a tendency toward future orientation in the camp group was found to be positively associated with the personality traits of pessimism and resentment. Future orientation was also found to be negatively associated with the sum of the ratings on the three positive personality traits. These findings suggest the need for a re-examination of the hypothesis that future orientation in the fantasies of 11-year-old children is a sign of a mature and healthy adjustment pattern. It seems, to the contrary, that the better adjusted children tended to be more present-oriented than those judged as less well-adjusted emotionally.

Further evidence in support of this revised hypothesis can be derived from the relation between friendship patterns and story scores in the disturbed sample. Thompson and Horrocks (5, 7) demonstrated developmentally with normal children that there is increasing stability of friendships with increasing chronological age. Thus, low friendship fluctuation scores may be considered an index of maturity. In the Bradley sample, a considerable degree of association was found between the friendship fluctuation scores and the story-completion scores, indicating that the children with the more stable friendships tended to be more present-oriented in fantasy.

In view of the evidence at hand, some tentative conclusions are appropriate with regard to time orientation in imaginal processes. A high degree of future orientation in 11-year-old children appears to be associated with inadequate personal and social adjustment. It seems plausible that a high degree of future orientation in fantasy indicates that the child is dissatisfied in the present and is investing a good deal of his energy in fantasy about the future. In other words, children who find their everyday lives unsatisfying and who try to escape their present situation by means of fantasies about the future may well be more disturbed, while the more stable children invest more of their energy and thoughts in present life situations. In considering these possibilities, it should be emphasized that what may be a sign of good adjustment at one age level may well be abnormal at another. For example, the normal 17-year-olds in Barndt and Johnson's study were more future-oriented than any of our groups of 11-year-old boys, whereas the 17-year-old delinquent boys were similar to our 11-year-olds in their present orientation. Understandably, it may not be until early adolescence that a person is normally concerned more with the future than with the present. Thus, although present-oriented at 11 years or younger may signify healthy adjustment, future orientation at some later age may reflect maturity and healthy emotional adjustment. These notions are largely speculations, but further research can certainly provide more definite answers.

The findings in regard to friendship patterns were unequivocal, providing quite convincing evidence that emotionally disturbed children tend to vary their friends over short intervals of time to a greater degree than do normal children. These significant differences in friendship fluctuations were found when the present group of disturbed children was compared both with the present normal groups and with the normal children studied previously by Thompson and Horrocks. Therefore, in addition to confirming an experimental hypothesis, the finding of striking similarities between the present normal *S*s and those studied in an independent investigation serves as a replication. As we (3) and others have argued elsewhere, there is great need for the worthwhile contributions of exact replications in psychological experiments. When the replications yield similar results, then the findings obtained in modifications and extensions of the original studies become all the more meaningful. In this regard, the previous work with normals has been extended to the social relations of disturbed children.

In the camp sample, there was a significant correlation between the ratings on emotional disturbance and friendship fluctuations scores, with the more disturbed children showing a greater change in friends. While there was no association between the clinical ratings and the friendship fluctuations in the Bradley sample, this may be attributable to the fact that these boys formed a very homogeneous group in terms of emotional disturbance. The fact that they were all institutionalized because of this difficulty indicates that they all were far out on the continuum of maladjustment, and this homogeneity would be conducive to low correlations between relative standing on this variable and scores on the measure of friendship fluctuations.

At any rate, the objective evidence is in keeping with clinical impressions of unstable social relations in people who suffer from severe psychopathology. In many cases of childhood emotional disorders, the parent-child relation is unstable and unsatisfying, and it is understandable that this unhappy interpersonal pattern may well generalize into a general distrust and other negative traits that make it difficult for the youngster to establish effective relations with either peers or adults. Thus, a vicious cycle is set up, with emotional disturbance leading to disturbed social relations, and the lack of social adjustment making it impossible for the disturbed child to obtain the kinds of gratification (e.g., social approval, lasting friendships, love) that are necessary for healthy personality development. A fundamental aspect of the psychotherapeutic process with children is the establishment of a trusting, secure, warm relation with a significant adult in the hope that on this foundation a good personal and social adjustment may be built. From both a therapeutic and a scientific point of view, however, many lines of research remain to be investigated before any adequate understanding is available of the interpersonal relations that characterize emotionally disturbed children.

SUMMARY

Three groups of 11-year-old boys were studied in the following settings: (*a*) residential psychiatric treatment center, (*b*) summer camp, and (*c*) public school. A story-completion technique was used to investigate time orientation as revealed in imaginal processes. Assessment of fluctuations in friendship patterns over a three-week period was the method used to study interpersonal relations. Ratings on general emotional adjustment and several positive and negative personality traits were obtained from qualified staff personnel.

In comparison with normals, as predicted, emotionally disturbed children showed more

unstable friendship patterns. However, contrary to prediction, their story completions were not more present-oriented. In the normal camp group there was a significant positive association between ratings on emotional maladjustment and degree of friendship fluctuation. In the disturbed group, the more present-oriented a child was on the fantasy measure, the more stable were his friendships. With normal children there was no association between time orientation and the measure of interpersonal relations.

The present *S*s' scores were compared with those obtained by Barndt and Johnson, using the identical story-completion technique, in a study of 17-year-old delinquents and nondelinquents. All of our groups were significantly more present-oriented than were their nondelinquents, while there were no differences between our groups and their delinquents.

No differences were found between friendship fluctuation scores of the present normal groups and scores obtained previously by Thompson and Horrocks in studies of normal urban children. In each of our groups, there was approximately 50% change in choice of best friend over a two-week period, a result which replicates theirs.

It was concluded that disturbed and normal 11-year-old boys do not differ in fantasy measures of time orientation, nor do they differ from older delinquents, but they are more present-oriented than older normal boys. Also, unstable social relations are characteristic of emotionally maladjusted children.

REFERENCES

1. BARNDT, R. J., & JOHNSON, D. M. Time orientation in delinquents. *J. abnorm. soc. Psychol.*, 1955, **51,** 343–345.
2. DAVIDS, A. Alienation, social apperception, and ego structure. *J. consult. Psychol.*, 1955, **19,** 21–27.
3. DAVIDS, A. The influence of ego-involvement on relations between authoritarianism and intolerance of ambiguity. *J. consult. Psychol.*, 1956, **20,** 179–184.
4. GOLDFARB, W. Psychological privation in infancy and subsequent adjustment. *Amer. J. Orthopsychiat.*, 1945, **15,** 247–255.
5. HORROCKS, J. E., & THOMPSON, G. G. A study of friendship fluctuations of rural boys and girls. *J. genet. Psychol.*, 1946, **69,** 189–198.
6. LESHAN, L. L. Time orientation and social class. *J. abnorm. soc. Psychol.*, 1952, **47,** 589–592.
7. THOMPSON, G. G., & HORROCKS, J. E. A study of friendship fluctuations of urban boys and girls. *J. genet. Psychol.*, 1947, **70,** 53–63.
8. WALLACE, M. Future time perspective in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **52,** 240–245.

Received August 7, 1957.

THE INTERACTION OF ACHIEVEMENT MOTIVATION AND ABILITY IN PROBLEM-SOLVING SUCCESS[1]

ELIZABETH G. FRENCH
Air Force Personnel and Training Research Center

THE main purpose of the present study was to test an hypothesis concerning the relationship between ability level, achievement motivation, and problem-solving performance. The experiment was also designed so that a previously established relationship (3) between achievement motivation and performance could be tested under somewhat different conditions. and the nature of that relationship could be examined further.

The present hypothesis, an outgrowth of earlier studies (1, 3), states that performance in a problem-solving situation will be related to ability level among subjects (*S*s) with high achievement motivation, but there will be little or no relationship among *S*s with low achievement motivation. The reasoning behind this prediction is that *S*s with high achievement motivation presumably will attempt to use their ability as effectively as possible, thus differences in that ability will be reflected in performance. *S*s with low motivation, on the other hand, will tend not to use their potential, so those with high ability will do little or no better than those with low ability.

The additional relationships studied concern the effect of achievement motivation on performance. A previous study (3) demonstrated that motivation was related to successful performance in a problem-solving situation where the length of time worked was determined by *S*. Although it could be shown that this relationship remained when length of time worked was partialled out, the present experiment was designed to test the hypothesis that the motivation-performance relationship exists when time worked is the same for all *S*s.

In the earlier study, also, it appeared that the relationship between motivation and problem success was not uniform over the entire motivation range. There was a steady increase in percentage of successes with increase in motivation interval up to about the median of the motivation distribution. Beyond that point the percentage of successes was roughly the same for all motivation scores. That is, any motivation level above the median was sufficient to give the maximum effect of that variable. This finding was not predicted nor expected on any theoretical grounds. The explanation could be that this is the general nature of the relation between motivation and performance or it could lie in the specific problem—its difficulty level, intrinsic interest value, etc. This experiment included in the design the possibility of testing for the existence of the described relationship under conditions slightly different from those where it was previously noted.

The specific hypotheses to be tested are:

I. Intelligence will be more highly related to success in problem-solving among *S*s with achievement motivation scores above the median than among *S*s with motivation scores below the median.

And the additional hypotheses:

II. *S*s with achievement motivation scores above the median will solve more problems in a given time period than will *S*s with motivation scores below the median.

III. Differences in motivation scores below the median of the distribution will be more highly related to problem-solving performance than will differences in motivation score above the median.

METHOD

Measure of Motivation

The measure of motivation was a projective device called the Test of Insight (2) in which the *S* is asked to provide "explanations" for briefly described behaviors presented as typical of hypothetical individuals. The number of achievement-relevant explanations make up the score.

The Problem Apparatus

The problem-solving apparatus consisted of a panel containing a horizontal row of seven lights and another of seven switches. The wiring of the board was such that some switches operated one light; some, two; and some, three. Similarly, some lights were controlled by one switch only and some by two or three. Operation of a switch always affected all the lights to which it was connected. Thus, operation of Switch 1, controlling Lights 1 and 3 would put those lights on; subsequent operation of Switch 2, controlling Lights 1 and 2 would extinguish Light 1 (already on) and put on Light 2. The *S*s were required to create specified patterns of lights.

---

[1] This report is based on work done under ARDC Project No. 7719, Task No. 17011, in support of the research and development program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

Subjects

The *S*s were 96 airmen undergoing basic training at Lackland Air Force Base. They were selected to cover the Armed Forces Qualifications Test (AFQT) range of 70–99, with equal numbers in the 70–79, 80–89, 90–99 ranges.

The *S*s were also selected so that equal numbers of each quartile of the achievement motivation fell in each AFQT decile. This made a 3 × 4 design with three levels of AFQT score and four levels of achievement motivation with eight *S*s per cell.

Procedure

Each day's *S*s were met as a group by *E* and given an orientation to the problems which established them as measures of general intelligence and all-round capability which gave an accurate estimation of what a man could accomplish. Immediately following this, they were given the Test of Insight, described as unrelated to the problems. Upon completion of this test, each *S* went individually to the problem-solving session.

First, *E* demonstrated the apparatus to *S* and stressed the important points, i.e., that switches could operate more than one light, and vice versa, and that the operation of a switch affected all the lights it controlled. Then *S* was permitted to operate the switches, one at a time, in order to become familiar with the switch-light combinations. At the end of about a two-minute practice period he was asked to recite these combinations to *E*. When *S* had demonstrated that he knew the combinations, *E* read off four demonstration problems together with their solutions, and *S* operated the designated switches to produce the light pattern stipulated. Then *E* read the following instructions:

> The problems you will solve will be like these you have just done. In working them out you may use the switches as often as you like but you may have only one on at a time, as you did when you were learning the combinations. In other words, you must work the solution out in your head, not on the board. When you think you have the answer, when you think you know which switches will give you the pattern of lights you want, tell me what you think it is so I can record it, and then try it out on the board the way you did the demonstration problems. If it is correct, if you get the correct lights, turn them off and go directly to the next problem. If it is wrong, turn the lights off and keep working on the same problem. Take each problem in turn, solving one before you go to the next. You will have 25 minutes to solve as many problems as you can.
>
> Is it clear? You work out the solution, using the switches one at a time if you like, call out all solutions before you try them, try the solution, if it works go to the next problem, if it doesn't keep working on the same one.
>
> The problems are written on these cards, one problem to a card. They tell you the number of switches you need to use and the lights you are to get on.

The problems were written on individual 3 × 5 cards and placed on the table by the apparatus. (An example: What *three* switches will give you Lights 1, 3, and 6?) The first four were warm-up problems designed to be solved in about one minute. Their purpose was to provide more familiarity with the apparatus and to minimize practice effect in the series proper so that mean solution times would be approximately the same for all problems. The regular series consisted of 16 equal-difficulty problems. It was anticipated that no *S* could complete the series in the time allotted. Scoring and timing started with the regular series, and *S* was given a two-minute warning of the time limit. The score was the number of problems solved in the 25-minute time limit.

RESULTS AND DISCUSSION

Before the main data analyses were carried out, the time spent on the preliminary problems with respect to its relation to number of regular problems solved and the individual problem mean times were examined. The mean time required to solve the preliminary problems was 5.69 minutes. This time varied widely within cells and also varied as a function of motivation score ($F = 3.67$, $P = .05$). Since the time spent on the preliminary problems varied in the same way as time per problem in the regular series (r for preliminary problem time vs. number of regular problems solved $= -.23$), it cannot be argued that the success on the regular series was due to a longer practice time in the preliminary period. In the regular series, the mean times for the first eight problems (not enough *S*s completed more than eight to permit stable means for the remaining problems) were, in order: 2.92, 3.21, 2.63, 3.54, 2.85, 4.65, 3.57, 2.91. They are all fairly homogeneous with the exception of that for Problem 6, which apparently was more difficult than the others. The effect of this greater difficulty of that one problem will be discussed later.

The mean AFQT and achievement motivation scores were also checked for individual cells. No differences were found in motivation means across a given AFQT level or vice versa.

The problem-solving scores ranged from 1 to

12, with a mean of 5.67 and a standard deviation of 2.33. The means for the 12 cells are presented in Table 1. The four quartiles of the achievement motivation distribution are delimited by scores 0–4, 5–7, 8–10, and 11–20.

Since the first two hypotheses dealt with differences expected as a result of motivation scores above or below the median, the two top and the two bottom quartiles were combined for the analyses pertaining to these hypotheses. The summary of the analysis of variance is presented in Table 2. The *F* for the interaction between AFQT and achievement motivation score (the test of the main hypothesis) is significant at the .01 level. Tukey's gap test gave a difference of 1.04 between means as a significant gap. By this test, the means for the three AFQT levels in the high motivation group are significantly separated from each other; those in the low motivation group are not. The relationships between AFQT and performance for the two motivation groups are also described by correlations of .60 and .08 for high and low motivation respectively. The hypothesis that ability level will be related to problem-solving performance among *S*s with high achievement motivation and unrelated among *S*s whose achievement motivation was low is supported by the data.

The second hypothesis, that motivation is related to problem-solving success when time worked is constant was also clearly supported by the analysis. The *F* for achievement motivation is significant beyond the .001 level. The correlation between the variables is .58.

TABLE 1

MEAN PROBLEM-SOLVING SCORES BY ACHIEVEMENT MOTIVATION AND AFQT LEVEL

AFQT	Motivation Score			
	11+	8–10	5–7	4–
90–99	8.75	9.00	4.12	4.00
80–89	7.25	6.87	4.25	4.12
70–79	6.00	6.00	4.00	3.62

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF PROBLEM-SOLVING SCORES WITH TWO LEVELS OF MOTIVATION AND THREE LEVELS OF AFQT

Source	Sum of Squares	*df*	Mean Square	*F*	*P*
Motivation	260.04	1	260.04	120.39	< .001
AFQT	39.14	2	19.57	9.06	< .01
M × AFQT	29.65	2	14.83	6.86	< .01
Within (error)	194.50	90	2.16		

The support for the first two hypotheses is fairly unequivocal and needs little comment. Hypothesis III requires a little more discussion. Examination of the means in Table 1 shows that the data fail to give it support. An analysis of variance of the 12-cell table of means breaking down the variance due to motivation into that due to position with respect to the median, that due to quartile within halves of the distribution, and the interaction of the two, yields *F* ratios considerably less than one for both the quartile position and the interaction between quartile position and position with respect to median. Thus it is apparently not true that in general the relation between achievement motivation and problem-solving is closer below the median of the motivation distribution than above it. In this case, the median of the distribution provided a sharp dividing line for problem-solving scores. Few *S*s with motivation scores below the median did well and few with scores above did poorly. There was very little overlap between the distributions. But the actual score above or below the median made little difference. That is, those with very high scores did not do noticeably better than those with moderately high scores, nor did those with very low scores do more poorly than those whose scores were moderately low.

In an attempt to find a tentative explanation for the failure to demonstrate the kind of relationship that had appeared in the previous experiment the distributions of problem-solving scores were examined for the several achievement motivation groups. This examination revealed first that the difference between the combined high motivation and the combined low motivation groups was manifested immediately. All the high *S*s were successful through Problem 4 and only two failed Problem 5. It wasn't until Problem 6 was reached that an appreciable number of those *S*s—six—failed. The failures in the low groups began with the second problem. Coming now to the two low quartiles, it is clear that they do not differ through Problem 4. Seventeen third and 16 fourth quartile *S*s reached and passed that problem. But with Problem 5, the beginning of a difference is suggested. The number of successes is 13 and 7, respectively. The next problem, Problem 6, it will be recalled, had a longer mean solution time than any of the others suggesting a higher difficulty level. At this point there is an abrupt drop in the number of successes among the low quartile groups. (It is also the point where failures first appeared in any number in the high groups.) Only four of the twenty who reached it completed it. It should be pointed out here that the average time available to these 20 *S*s when they started Problem 6 was 5.44—one minute longer than the average solution time for that problem. No low quartile *S* completed more than six problems. The proportion of successes—3 to 1—still favors the third quartile, but the numbers are so small that they can be only suggestive.

After considering these points, we suggest the explanation that the greater difficulty of Problem 6 prevented the more gradual dropping out of third quartile *S*s as compared to fourth quartile *S*s, since it stopped nearly all low *S*s. Had that problem been comparable to the others in level of difficulty, the trend started with Problem 5 might have continued more gradually. The difficulty level of Problem 6 also probably contributed to the lack of overlap in problem scores between the two halves of the motivation distribution. The tentative conclusion is that the difficulty of the task is one factor in determining the kind of relationship which will appear between achievement motivation and successful completion. This is a point for further investigation.

In summary, it may be said that the main hypothesis was clearly supported. It seems evident that, in this kind of a situation at least, performance depends not only on having the ability to do well but on the motivation to use that ability as well. The form the relation between achievement motivation and performance takes appears to depend in part on the characteristics of the task.

SUMMARY

Three hypotheses were tested: that AFQT would be related to problem-solving performance among *S*s with high achievement motivation but not among *S*s with low achievement motivation; that motivation would be related to performance when all *S*s worked the same length of time; and the relation between motivation and performance would be greater below the median of the motivation distribution than above it.

Ninety-six *S*s, eight in each of the 12 cells of a 3 × 4 table representing three levels of intelligence and four levels of motivation score, solved as many problems as they could in 25 minutes on a switch-light pattern apparatus.

The results indicated clear support for the first two hypotheses but none for the third. There is some indication that the difficulty of the task influences the motivation-performance relationship so that this will not be constant from task to task.

REFERENCES

1. FRENCH, ELIZABETH G. Some characteristics of achievement motivation. *J. exp. Psychol.*, 1955, **50**, 232–236.
2. FRENCH, ELIZABETH G. *Development of a measure of complex motivation. USAF Pers. Train. Res. Cent., Res. Rep.* April 1956, No. AFPTRC-TN-56-48.
3. FRENCH, ELIZABETH G., & THOMAS, F. H. The relation of achievement motivation to problem-solving effectiveness. *J. abnorm. soc. Psychol.*, 1958, **56**, 45–48.

Received August 13, 1957.

FAILURE AND THE SUBSEQUENT PERFORMANCE OF SCHIZOPHRENICS[1]

GORDON W. OLSON[2]

State University of Iowa

To account for psychological deficit (9) in schizophrenia, the hypothesis is often advanced that schizophrenics are less responsive to environmental stimuli than are nonschizophrenics (5, 11). In direct contrast to this view, it has also been suggested that schizophrenics are withdrawn from social contacts because they are hypersensitive to social stimuli, notably censure (2). Supporting the hypothesis of hypersensitivity are the studies of Garmezy (6), Bleke (1), Webb (15), and Hahn (7). These investigators all reported deleterious effects on the performance of schizophrenics from social punishment. Cohen (3), on the other hand, found that a mild physical punishment (shock) was more successful in maintaining the response level of schizophrenics than a verbal "rapport" condition. These findings suggest a possibly important difference in the reaction of schizophrenics to mild social punishment and to mild physical punishment and lend weight to the idea of a specific hypersensitivity to social punishment in schizophrenia.

Studies of success and failure in normals typically yield results which appear to be directly opposite to those found for schizophrenics. Consistently, mild failure conditions appear to facilitate the performance of normal subjects to a greater extent than success conditions (4, 12, 13). Social rewards and social punishment seem, therefore, to produce directly opposite effects in schizophrenics and normals. In an attempt to clarify some of these issues, this study was designed to investigate the comparative effects of positive, negative, and nonevaluative verbal statements on the performance of schizophrenic and normal subjects.

METHOD

Task. The task was constructed with a view toward maximizing performance and minimizing new learning in an area involving verbal-symbolic performance. A digit symbol test was developed, requiring the substitution of a caret for the numbers 1, 2, 3, and 4, the apex of which rotated 90° clockwise for each consecutive number, i.e., 1 ∧, 2 >, 3 ∨, and 4 <. The measure used was the number of *attempted* items per 50-second trial. The number of attempted substitutions was used rather than the number correct because a pilot study revealed a correlation of +.95 for the two measures, and motivation, not efficiency, was the subject of main concern. The sequence of digits on the worksheet was derived from a table of random numbers, and when a number appeared twice in succession, the next occurring number replaced the duplicate. The substitution code appeared at the top of the worksheet. A maximum score of 100 was possible.

Subjects. The 45 schizophrenic *S*s were male patients at the Knoxville, Iowa, Veterans Administration Hospital between the ages of 20 and 40, with an educational achievement level of not less than the sixth grade. All patients carried an unchallenged diagnosis of schizophrenia which encompassed the usual subtypes. The length of hospitalization ranged from six months to 17 years, the mean being 4.3 years. The 45 normal *S*s were randomly selected male psychiatric aides at the same hospital who met the same age and educational criteria as the schizophrenics. No psychotic *S* had experienced psychosurgery or had received electric or insulin shock therapy within the previous two months. If *S* was presently on a tranquilizing drug, this fact was noted; the tranquilized *S*s seemed approximately equally distributed among the three instructional subgroups. A reasonable degree of homogeneity was sought by presenting each *S* with not more than two adaptation trials of 15 substitutions each prior to the experiment proper. Any *S* who failed to complete the second such trial with less than three errors was discontinued. Thus, the experimental *S*s were a select group of hospitalized schizophrenics with respect to cooperation and ability to learn a simple task.

Procedure. Prior to the first experimental trial, *S* learned the substitution code to two correct trials by orally reporting the direction of the apex when only the numeral was exposed on a card. The untimed adaptation trials of 15 items followed without reference to a cue sheet. Errors occurring during the adaptation trials were pointed out by *E* and corrected by *S* in order to further the learning process. Errorless performance on the first adaptation trial excused *S* from the second, and the experimental trials were immediately initiated.

For the experimental trials, the test sheets were drawn from a large pile within sight of *S*, who was not told how many trials were forthcoming. Each trial consisted of 50 seconds of work followed by a 10-second rest. Five such trials constituted a block, and the three blocks were separated by six-minute rest intervals during which *E* scored the preceding five trials. Positive, negative, or nonevaluative conditions were introduced during the last minute of the rest intervals between Blocks I and II and Blocks II and III, but not between trials within blocks. Assignment to the subgroups defined by the conditions was consecutive.

---

[1] Based upon a doctoral dissertation submitted to the State University of Iowa. The author wishes to express his gratitude to the VA Hospital, Knoxville, Iowa, for the cooperation necessary for conducting this study.

[2] Now at Anoka State Hospital, Anoka, Minnesota.

The rest interval statements were as follows, with only minor modifications to insure naturalness and conviction; the emphasis was on instituting the desired state, rather than presenting an exact sequence of words:

Positive: I have scored your papers, and I see that so far you are doing very well. As a matter of fact, you are doing much better than most men of your age and background. Let's try it again and see what happens.

Negative: I have scored your papers, and I see that so far you are not doing very well. As a matter of fact, you are doing very poorly compared with other men of your age and background. Let's try it again and see what happens.

Nonevaluative: I have scored your papers during the rest period which we just had, but we still have some more to do, so give me just a second or two to get the material rearranged and then let's do it again and see what happens.

Through Block I only two groups are differentiated, schizophrenic (S) and control (C). For Blocks II and III, six groups were designated, identified by appropriate symbols: schizophrenic, nonevaluative (SO); schizophrenic, positive (SP); schizophrenic, negative (SN); control, nonevaluative (CO); control, positive (CP); control, negative (CN). There was no predominance of any subtype of schizophrenia in any of these categories. Reassurance of adequate performance was given both the SN and CN groups after completion of the experiment.

RESULTS

Among the patients tested, six were unable or unwilling to participate at the required criterion level. Among the normal *S*s, all scheduled for participation took part with the exception of one who requested exclusion for personal reasons. The mean ages for the S and C groups were 29.4 years and 33.2 years, respectively, and the standard deviations were 5.21 and 3.95. While the normal group was significantly older ($p < .01$) than the patient group, the negative correlation of age and performance should attentuate rather than exaggerate any differences favoring the normals. The mean educational level for the schizophrenics was 10.9; for the normals it was 10.4. The standard deviations were 2.75 and 1.89, respectively.

The mean numbers of attempted substitutions for each subgroup are presented in Table 1. Analysis of the results for Block I indicates that the initial performance of the schizophrenic group was clearly inferior to that of the normals ($p < .01$), although both groups showed significant improvement over the five trials ($p < .01$). Analysis of variance also indicates that both diagnostic groups and all six subgroups improved over these successive trials at the same rate, and that there was no significant difference in performance over the initial trials among the subgroups by conditions.

Employing a Lindquist (10) Type III design, the analysis of variance of the two diagnostic groups, the three subgroups by conditions, and the three blocks again indicates superior performance for the normal group ($p < .01$). Significant improvement in performance across blocks is also demonstrated ($p < .01$), as is significant improvement in performance due to conditions ($p < .05$). None of the interactive effects were significant. Analysis of the effects of the conditions for Blocks II and III reveals a higher level of performance for *S*s under either the positive or the negative statements than those under the nonevaluative condition ($p < .01$). The positive and negative conditions did not appear to differ in their effectiveness.

Analysis of the data for the schizophrenics

TABLE 1

MEAN NUMBER OF ATTEMPTED SUBSTITUTIONS PER TRIAL

Group	Trials														
						Block II					Block III				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
CP	31.3	32.1	33.5	33.5	33.7	40.5	39.5	39.5	38.4	38.6	45.1	43.3	42.9	41.9	40.9
CN	28.0	30.7	32.1	32.5	32.6	40.6	43.3	44.3	44.3	43.6	48.6	48.9	48.1	48.5	48.2
CO	28.5	30.3	31.6	32.3	33.9	37.2	36.7	37.8	37.2	37.4	42.3	40.7	40.1	39.9	39.3
SP	24.2	26.6	26.9	28.3	29.3	34.6	35.5	36.7	36.8	37.0	41.7	40.1	40.5	40.7	40.1
SN	23.0	25.5	25.7	25.9	28.7	32.7	33.6	32.9	33.2	34.1	36.9	37.3	37.9	36.5	35.6
SO	23.7	25.3	27.3	27.7	29.1	31.7	30.7	30.9	30.2	30.9	35.0	34.5	32.2	32.7	30.5

only reveals an interaction between conditions and blocks ($p < .01$), indicating the differential influence of the statements on the subsequent performance of the schizophrenic *S*s. Both positive and negative conditions produced enhanced schizophrenic performance as compared with nonevaluation (in both cases, $p < .01$), and the positive statements resulted in significantly greater improvement than did the negative ($p < .01$).

Among the normals only, there was a significant interaction of instructions and blocks ($p < .01$). Study of the simple effects of conditions indicated that negative statements resulted in significantly improved performance as compared with the nonevaluative condition ($p < .01$), but the positive instructions apparently did not. Also, the negative condition for normal *S*s was followed by significantly better performance than was the case for positive statements ($p < .01$).

A comparison of the simple effects of conditions between diagnostic groups showed differential effects. In the two nonevaluative groups, the normals improved more over blocks than did the schizophrenics when neither positive nor negative statements were made ($p < .01$). The schizophrenics, however, performed better than the controls when both were under the positive condition ($p < .01$), and the controls showed significantly greater improvement than the schizophrenics when both were under the negative condition ($p < .01$).

It is of interest to examine briefly the relative amount of improvement shown by the subgroups as a function of rest and conditions. When improvement is defined as the amount of gain from the last prerest trials to the first postrest trial, there was an insignificant over-all difference in improvement scores in favor of the normals. The difference in improvement was significant ($p < .05$) when improvement after the first rest period was considered separately, but the effect was not present following the second rest period. Since the over-all improvement data was confounded with the effects of the conditions, comparison was made of the improvement scores for the two nonevaluative groups only. No difference in improvement scores after either rest was found.

DISCUSSION

The results of this study indicate that negative statements, while not harmful to the performance of schizophrenics, were less effective as a facilitator of performance than was praise; similarly, failure was a less effective enhancer of performance for schizophrenics than it was for normals. This interpretation is in marked contrast to the conclusions of other investigators (1, 6), who in similar studies underscored the detrimental effects of social punishment in schizophrenia. The point merits emphasis that the present conclusions are based on the comparisons under a nonevaluative condition, a condition not included in either of the studies by Garmezy or Bleke.

Consistency of results of the present study and the above studies is found in the differential performance of the normals and schizophrenics to the positive and negative conditions. In each study, the schizophrenic *S*s did less well under negative conditions than positive; and in both Garmezy's and the present study, the normals showed better performance under the negative statements than positive. Although the differential effect of praise and failure for the diagnostic groups falls short of significance in the present study ($p > .10$), this differential response appears in the increasing difference over trials under the negative condition and the decreasing difference over trials under the positive condition.

The theory involving hypersensitivity to failure in schizophrenia (1) calls for the obvious prediction of a greater gain in performance for the normal negative than for the schizophrenic negative group, and this prediction was borne out here. A second prediction from this theory is that the schizophrenic negative group would do less well than the schizophrenic positive group; this difference, too, was found. Thus, the hypothesis of the deleterious effect of social punishment *relative to social reward* in schizophrenia seems substantiated. That the deleterious effect of punishment is only *relative* to the positive condition is clearly shown by the significantly greater improvement of the schizophrenic negative group over the schizophrenic nonevaluative group, in that if social punishment were absolutely deleterious, it would be necessary that the schizophrenic negative group do less well than the schizophrenic nonevaluative group.

It could be suggested that the initial deficit shown by the schizophrenics was due to a deficit in motivation, except that the significant difference in Block III still favored the normals. Thus, the initial diagnostic difference was not overcome by enhancement reflecting possible increments in motivation. Nevertheless, the enhanced performance of schizophrenics on this task equalled the unenhanced performance of normals, suggesting that motivational factors may have compensated for the diagnostic difference.

The finding of an increasing difference between the nonevaluative groups is interesting with respect to the concept of a differential rate of accumulation of inhibition for schizophrenics and normals (14). Support for such a notion is found in the increasing difference between the diagnostic groups under the nonevaluative condition, but refutation is found in the absence of differences in performance by the groups over trials within blocks. After this hypothesis, and following the paradigm of Hull (8) for reactive inhibition, it would be expected that the schizophrenic group would show more decline over the trials within blocks than would the normals. In addition, the scores for improvement after rest show that there was a greater amount of improvement for the normals, a finding in opposition to an interpretation of a greater accumulation of reactive inhibition for the schizophrenic *S*s.

SUMMARY

The present study investigated the questions of whether deficit would obtain for schizophrenics on a timed task of nonsocial content and, if so, whether the deficit would increase or decrease over trials under a condition of nonreinforcement and, finally, whether success or failure instructions would differentially affect the subsequent performance of schizophrenics and normals.

Forty-five schizophrenic males and 45 normal males were administered 15 timed trials of a simple digit-symbol task and the number of attempted substitutions per trial recorded. During the two rest periods which followed each block of five trials, the *S*s were consecutively assigned to positive, negative, and nonevaluative groups and informed respectively that they were doing very well, very poorly, or they were administered no evaluative comments relative to their immediately prior performance.

The results indicated that both schizophrenics and normals responded with enhanced performance to the negative conditions, but that only the schizophrenics demonstrated enhanced performance under the positive condition. There is suggestive evidence that the positive and negative statements affected schizophrenics and normals in a differential manner, with schizophrenics benefiting more from the positive than from the negative, but there is no evidence to suggest that the negative condition had a *detrimental* effect upon schizophrenics. The data are consistent with other studies reporting differential effects of social punishment and social reward for schizophrenics and normals. The data do not support the concept of increased reactive inhibition in schizophrenia, when the Hullian paradigm for reactive inhibition is employed.

REFERENCES

1. BLEKE, R. Reward and punishment as determiners of reminiscence effects in schizophrenic and normal subjects. *J. Pers.*, 1955, **23,** 479–498.
2. CAMERON, N. Reasoning, regression, and communication in schizophrenics. *Psychol. Monogr.*, 1938, **50,** 1–34.
3. COHEN, B. D. Motivation and performance in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **52,** 186–190.
4. FARBER, I. E., RUSSELL, W. A., & ANDREAS, B. G. Effect of failure upon performance in verbal and motor learning situations. *Amer. Psychologist*, 1949, **4,** 351. (Abstract)
5. FENICHEL, O. *The psychoanalytic theory of neuroses.* New York: Norton, 1945.
6. GARMEZY, N. Stimulus differentiation by schizophrenic and normal subjects under conditions of reward and punishment. *J. Pers.*, 1952, **20,** 253–276.
7. HAHN, I. N. The effects of success and failure experiences on chronic schizophrenic patients. Paper read at Amer. Psychol. Ass., Chicago, 1956.
8. HULL, C. L. *Principles of behavior.* New York: Appelton-Century, 1943.
9. HUNT, J. McV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. II. New York: Ronald, 1944. Pp. 971–1032.

10. LINDQUIST, E. F. *The design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
11. PIOTROWSKI, Z. Work habits as a means of differentiating native and acquired deficiency. *Psychol. Bull.*, 1933, **30**, 728–729.
12. RUSSELL, W. A., & FARBER, I. E. Retention of verbal material as a function of degree of failure experienced in original learning. *Amer. Psychologist*, 1948, **3**, 355. (Abstract)
13. STEISEL, I. M., & COHEN, B. D. The effects of two degrees of failure on level of aspiration and performance. *J. abnorm. soc. Psychol.*, 1951, **46**, 79–82.
14. VENABLES, P. H., & TIZARD, J. Performance of functional psychotics on a repetitive task. *J. abnorm. soc. Psychol.*, 1956, **53**, 23–26.
15. WEBB, W. W. Conceptual ability of schizophrenics as a function of threat of failure. *J. abnorm. soc. Psychol.*, 1955, **50**, 221–224.

Received August 16, 1957.

THE EFFECTS OF PLAUSIBILITY ON THE LEARNING OF CONTROVERSIAL STATEMENTS[1]

EDWARD E. JONES AND RIKA KOHLER
Duke University

PERHAPS one of the most ubiquitous assumptions of recent research on human cognition is that perception, learning, and memory operate in ways which are instrumental to adjustment and survival. This assumption has given rise to a number of experiments in the tradition of perception research which have shown, for example, that our cognitive activities tend to be selective (12, 19), that we accentuate (9) or fixate (17) stimuli that are symbolic of reward, and we fail to perceive (13) or autistically distort (15) stimuli that are inimical to our own values or self concepts. These phenomena all seem to exemplify the trend toward cognitive constancy or stability of attitude.

While many of these studies have been criticized on both procedural and logical grounds (e.g., 14, 16), there seems less resistance to the contention that the so-called higher mental processes of learning and association are conditioned by adjustive requirements. The most clear-cut demonstrations of such effects have occurred in the area of meaningful learning, where numerous attempts have been made to show that we learn best those statements which are congenial to our beliefs and attitudes (e.g., 4, 10, 18, 20). The modest but fairly consistent success of these attempts has given support to those impressed by the autistic nature of learning. The proposition that we learn best that with which we agree was originally derived from basic considerations of the functional or instrumental role of learning. The generality of this proposition depends, however, on the extent to which it really is adjustive to be especially retentive of congenial views.

The fact that we are, under normal circumstances, relatively unreceptive to facts and arguments that run counter to our biases would appear to be a potent barrier to attitude change. To the extent that autism is operative, the recipient of an influence communication will have difficulty learning the communication message and/or will distort the content of the message so that it conforms more closely to the position he already holds.

It is fortunate for the experimenter, the propagandist, or the educator who attempts to modify a significant attitude that all learning and retention is not autistic, at least in the more conventional sense of the word. In the first place, the very fact of exposure is certainly important. An individual may retain more of a co-valuant than a contra-valuant communication, but he still may retain significant portions of the latter. Furthermore, it is quite easy to create conditions in which a reversal of the simple autism effect occurs. Jones and Aneshansel (8) have shown how prosegregation *S*s learn antisegregation statements *better* than antisegregation *S*s if both groups anticipate a task of providing appropriate counter-arguments to prosegregation statements. When there is no mention of any counter-argument task, the usual autism effect demonstrated most clearly by Levine and Murphy (10) is observed. Thus, *S*s may be motivated by an ulterior incentive to learn well those statements with which they disagree, a finding which bears a relation to studies of role-playing in attitude change research (cf. 7).

Other considerations have to do with the plausibility or extremity of the arguments involved in the communication. It seems reasonable to assume, as Hovland, Janis, and Kelley (5) have done, that "a communication only slightly removed from the individual's position will sometimes influence him positively while one far removed will affect him negatively (boomerang effect)"[2] (p. 286). What is the significance of this assumption for the learning process itself? Obviously, the extreme message must be assimilated in some way in order for a boomerang effect to occur. Beyond this, however, the individual may be positively motivated to absorb and fixate the extreme statement. He may be avidly receptive to ludicrous overstatements of positions differing from his own, just as the arch conservative tends to fixate in his memory statements made by liberal associates which may be construed as "socialistic." One's own self-righteousness is enhanced by indications of the implausibility of the opposition; therefore, such indications will be retained and cherished as self-supportive.

---

[1] This study was conducted in the program of the Organization Research Group of the Institute for Research in Social Science at the University of North Carolina. The O.R.G. is supported jointly by the Ford Foundation and the Office of Naval Research [Nonr-855 (04)]. The research was partially financed by a grant to the senior author from the Duke University Research Council.

[2] A recent study by Hovland and Pritzker (6) indicates that this does *not* happen when the issue is relatively uninvolving and the prestige of the source is high.

The converse is also true. We are embarrassed when others give ineffective support to our own cause, just as the educated Southerner may be embarrassed by the Georgia cracker. As a consequence, we rush to forget the false and implausible arguments that favor our position but make salient the negative elements in our own reference group. If these common sense deductions are true, then we would have to qualify further the statement that people learn best those statements which are in favor of their position. In order for the conventional autism effect to hold, the statement would have to be sufficiently plausible and effective so that the individual could comfortably endorse both the position implied by the statement and the supporting reasons adduced.

The present experiment is concerned with the effects of both plausibility and direction on the learning of controversial statements. In brief outline, we propose to examine the proposition that *S*s who hold to one set of beliefs learn plausible statements in favor of those beliefs and implausible statements opposing them better than they learn implausible-favoring or plausible-opposing statements. In more specific terms, three related hypotheses are tested:

1. Prosegregation *S*s learn plausible prosegregation and implausible antisegregation statements with less difficulty than they learn plausible antisegregation and implausible prosegregation statements.
2. The reverse is true of the antisegregation *S*s.
3. Those with neutral or intermediate beliefs learn all types of statements equally well.

A preliminary experiment and a more rigorous replication experiment were conducted to test these hypotheses.

EXPERIMENT I

Method

Subjects. A questionnaire obviously dealing with attitudes toward segregation was administered to 299 introductory psychology students. One set of items on this questionnaire constituted a self-rating scale which the students could check from "1. Strong prosegregation. Example: I would favor all measures short of actual violence to insure the long-run maintenance of segregation" to "9. Strong prointegration. Example: I would favor relying on the full power of Federal and State authorities to insure that desegregation occurs as rapidly as possible." After indicating their position "in general" on this nine-point scale, the students were confronted with a number of community situations that would presumably make integration difficult to effect (e.g., "There are many more Negro school-age children than white, and almost as many Negro teachers as white teachers"). For each situation, the students were to check a number from one to nine, indicating their action attitude as a responsible member of the community described. Their initial self-rating was averaged with self-ratings repeated across five community situations to obtain a final attitude score. From the pool of 299 students, 45 *S*s were selected: 15 prosegregation *S*s (hereafter referred to as "pros") with scores ranging from 1.11 to 2.50, 15 intermediate *S*s (hereafter referred to as "neutrals") with scores ranging from 3.50 to 4.50, and 15 antisegregation *S*s (hereafter called "antis") with scores ranging from 6.85 to 7.98.

Experimental procedure. The actual procedure involved in the learning task was very similar to that reported in (8). Each of the 45 *S*s attempted to learn 12 statements in individual sessions. The statements were printed on separate cards and each *S* was instructed to read the cards aloud at normal reading speed. After reading the 12 statements in this manner, *S* was asked to reproduce as many statements or phrases as he could remember. This identical procedure was repeated for five learning trials. The order of statements differed randomly on each trial, but the order was constant across *S*s.

Six of the 12 statements to be learned were prosegregation in direction, six were antisegregation. Within each group of six statements, three were "plausible" and three were "implausible." Plausibility was determined by the pooled rankings of eight graduate student judges. Each judge was presented with 20 statements, 10 of which were obviously prosegregation and 10 of which were obviously antisegregation. Out of each group of 10 statements, the three which were ranked lowest and the three which were ranked highest on "plausibility" were selected for use in the experiment. All statements were equated for length. Examples of each type of statement follow:

1. *Plausible prosegregation.* Southerners will have to pay the price of lowered scholastic standards if they yield to the pressures to integrate their schools.
2. *Implausible prosegregation.* If Negroes and whites were meant to live together, they never would have been separated at the beginning of history.
3. *Plausible antisegregation.* The present inferior condition of the Negro is the result of long and effective suppression by the Southern whites.

4. *Implausible antisegregation.* The real reason why most Southern whites oppose integregation is their realization that the Negro is more capable than they are.

While there is room for debate concerning the sense in which a given statement is plausible or implausible, it was important to avoid statements which were completely ridiculous as well as those which were completely truistic. The judges, thinking in terms of the experimental *S*s that were to be used, showed a high degree of agreement in their rankings of statements.

Results

Each reproduction attempt was given a score of 1 if it expressed a vague idea of part but not all of the statement; 2 if it expressed the over-all idea of the entire statement; 3 if it was correct except for the use of synonyms; and 4 if it was verbatim. Because of the greater possibility on the earlier trials of reproductions that combined statements or fell between statements, and the consequent unreliability of scoring, analysis was restricted to reproductions made on Trial 5.

The results of Experiment I provide rather firm support for the hypotheses. As Table 1 indicates, the appropriate comparisons between pro and anti *S*s on each of the four statement types are in the right direction in every case, and three of the four differences are significant at the .05 level or better. In every case but one, the neutral *S*s fall between the pro and the anti *S*s on their fifth trial learning scores. It would appear that plausibility, in combination with the direction of a controversial statement, is a variable of some importance in predicting learning success. The main reversal of prediction is accounted for by the antis' relatively poor learning of the plausible anti statements and their relatively good learning of the implausible anti statements.

In spite of the generally good support for the experimental hypotheses, there are a number of procedural criticisms that might be directed at this first experiment. The reliability and validity of the attitude self-ratings are unknown, *E* may have unwittingly influenced the results by on-the-spot scoring, the sensitivity and reliability of scoring procedures which consider only the final trial might be considered suspect, and the generalization value of the results is contingent on the representativeness of the particular statements chosen as plausible and implausible. For these and other reasons a replication was designed.

TABLE 1

MEAN LEARNING SCORES FOR TRIAL 5 BY ATTITUDE AND STATEMENT TYPE

(Experiment I)

Subjects	Statements			
	Pro		Anti	
	Plausible	Implausible	Plausible	Implausible
Pro	2.7	1.4	1.3	2.6
Neutral	2.1	2.0	2.4	2.3
Anti	1.5	2.6	2.2	2.1
Anti vs. Pro P_{diff}[a]	<.01	<.01	<.05	—

[a] By *t* test, two-tailed.

EXPERIMENT II

Method

The learning procedure itself was identical to that of the first experiment except that *E* did not know the *S*'s attitude category until the session was over, and each reproduction was transcribed verbatim on specially designed recording sheets. The major changes occurred in the following respects.

Subjects. Once again introductory psychology students served as *S*s, but they were selected by a different attitude measure. A 10-item Likert-type scale was administered to 147 students. This scale was identical to that used by Jones and Aneshansel (8), and its characteristics and validity are described in the earlier article. In the present study, scores ranged from a low of 13 (possible minimum, 10) to a high of 66 (possible maximum, 70) and the distribution was somewhat skewed toward the low, or antisegregation, end of the scale. Thirty-nine *S*s were selected: 13 pros with scores ranging from 43 to 66, 13 neutrals with scores from 28 to 34, and 13 antis with scores from 13 to 21.

Statements. Twelve statements, all of which differed from those used in the first experiment, were used in the learning task. Once again, three statements were pro-plausible, three anti-plausible, three pro-implausible, and three anti-implausible. The same procedure was followed to determine statement plausibility, namely the pooled rankings of, this time, six judges.

Scoring. Each statement was divided into four items, which represented single idea units. An item was then scored as correct if it was reproduced verbatim or only incidentally modified by the use of synonyms. As a check on the reliability of scoring, 1459 items were systematically selected to sample equally from the 39 *S*s and the 12 statements. These items were independently scored by a second rater and there was 90% agreement on categorization of "correct" items.

Results

The results of the second experiment are even more dramatic than those of the first. The findings are graphically summarized in Fig. 1. Here it may be seen that the hypotheses are confirmed in every detail, and for each type of

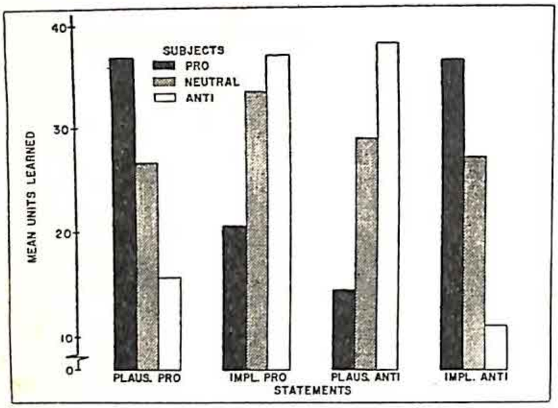

FIG. 1. THE RELATION BETWEEN PLAUSIBILITY, DIRECTION OF STATEMENT, AND ATTITUDE IN THE LEARNING OF CONTROVERSIAL MATERIAL

TABLE 2

ANALYSIS OF VARIANCE OF LEARNING SCORES POOLED OVER FIVE TRIALS

(Experiment II)

Source	df	Mean Square	Error Term	F
A: Attitude	2	105.63	b	<1.00
Error b	36	274.03		
Between Ss	38			
B: Direction	1	210.01	W_1	5.18*
C: Plausibility	1	16.68	W_2	<1.00
A × B	2	22.51	W_1	<1.00
A × C	2	148.86	W_2	6.17**
B × C	1	354.00	W_3	3.24
A × B × C	2	5902.63	W_3	53.99***
Error W	108	58.00		
Error W_1	36	40.53		
Error W_2	36	24.13		
Error W_3	36	109.33		
Within Ss	117			
Total	155			

* $p < .05$.
** $p < .01$.
*** $p < .001$.

statement the neutral *S*s fall between the pros and the antis. Thus the familiar Levine and Murphy effect holds only for plausible statements. As far as implausible statements are concerned, *S*s tend to learn better those statements which are opposite in direction to their own position.

The major statistical analysis of these results involved assigning each *S* a score representing the number of items correctly reproduced for each type of statement on each of the five learning trials. In order first to determine the effect of trials, these scores were placed in a mixed factorial design similar to that described by Edwards (3) for repeated measures on the same *S*s. Using the pooled within-*S*s error term advocated by Edwards, the main effect of trials was highly significant ($p < .001$), but none of the interactions *involving trials* approached significant values. Therefore, items learned to the criterion were summed over all five trials for each *S*, and the resulting scores were analyzed in a Type VI mixed design (following Lindquist, [11]). Collapsing over trials in this manner made it possible to use more conservative, and more precisely appropriate, error terms for each statistical effect without making the design impossibly cumbersome.

The results of this major analysis are presented in Table 2. In statistical terms, our hypotheses combine into the single prediction of a significant interaction between attitude, statement direction (pro or anti), and plausibility. With 2 and 36 degrees of freedom, an *F* ratio of 8.42 would be significant at the .001 level. The obtained *F* ratio for this interaction of 53.99 is far beyond the range of conventionally tabled probability values and is clearly significant in the extreme.

There are two other significant sources of variation, ignoring the above mentioned main effect for trials which, as plotting the means shows, is a function of the increase in number of items correctly reported from trial to trial. The significant main effect of direction reflects the fact that slightly more pro than anti statements are learned overall. The significant interaction between attitude and plausibility tells us that in the present experiment antis tended to learn plausible statements better, while pros tended to learn implausible statements better.

In view of the design of the experiment and the fact that each of these significant effects is dwarfed in magnitude by the higher-order interaction, the generalization value of these unpredicted differences would probably be limited in any event. In addition, none of these significant effects (except that of trials and the predicted higher-order interaction) replicates findings of the first experiment. In fact, more anti than pro statements were learned overall

TABLE 3
COMPARISONS BETWEEN ATTITUDE GROUPS FOR DIFFERENT STATEMENT TYPES
(Probability values for *t* test results, Experiment II)

Statement Type							
Pro Plausible		Pro Implausible		Anti Plausible		Anti Implausible	
Direction	*P*	Direction	*P*	Direction	*P*	Direction	*P*
P > N	.05	A > N	.50	A > N	.10	P > N	.05
N > A	.10	N > P	.01	N > P	.01	N > A	.001
P > A	.001	A > P	.01	A > P	.001	P > A	.001

Note.—P, N, and A stand for pro *S*s, neutral *S*s, and anti *S*s respectively.

in that experiment. In addition, the anti *S*s actually learned more implausible and fewer plausible statements than did the pro *S*s.

It is important to note that the *S*'s attitude bears no relation to the total amount of learning that occurred. This finding removes any reason for concern with the differential intelligence or learning ability of pro and anti *S*s in the present experiment.

While the analysis of variance is extremely useful in correcting for the lack of independence between the probability values of individual paired comparisons, the interaction between attitude, direction, and plausibility is impossible to interpret without reference to Fig. 1. To facilitate interpretation further, probability values of *t*-test comparisons between the means of different subject groups are presented in Table 3. From this table it becomes clear that the antis not only differ from the pros at the .01 level for every comparison; antis and pros in turn differ significantly from the neutral *S*s on five out of eight comparisons.

DISCUSSION

Both experiments provide strong support for the experimental hypotheses, which concern the importance of plausibility in the learning of controversial statements. When *S* is confronted with plausible statements that could be effectively used either in support of or in opposition to his own position, he tends to learn best those statements which argue in the direction of his own attitudinal bias. When the statements are overdrawn, implausible, or based on fallacious premises, however, *S* tends to learn best those statements which argue *against* his own attitudinal bias. Neither plausibility per se, nor the direction of the statement per se, seems to affect the learning process (cf. 20).

Those who favor a functionalistic view of cognitive processes will have little difficulty assimilating the results of the present experiment. Working from the assumption that *S* is concerned with protecting his attitudinal position, we see him ignoring or responding defensively to the effective or plausible opposing argument and reacting with almost hypervigilance to the implausible opposing argument. It is usually assumed that the greater the threat to the organism, the more likely will vigilance predominate over defense (e.g., 2). This assumption would imply that the implausible opposing argument is more threatening than the plausible, an unlikely state of affairs in any context simulating a debate in which rationality is rewarded. Furthermore, such an assumption leaves it very difficult to explain why *S*s have so much difficulty in learning the implausible statement that favors their own position. The explanation that we prefer is that the protecting of an attitudinal position is furthered by the denial both of implausible elements in one's own belief and plausible elements in opposing beliefs. Correspondingly, there is a tendency to fixate plausible favoring and implausible opposing beliefs.

The magnitude of the present effect is so striking that it prompts consideration of the possibilities of such a learning task for application to the assessment of attitudes. Particularly when one is dealing with attitudes that are opposed to cultural values and susceptible to dissimulation, it may be worth while to consider the joint manipulation of plausibility and direction in a learning task. Looking once again at the data of the second experiment, if we devise an index for each *S* that represents the difference between pro plausible and anti implausible on the one hand and anti plausible and pro implausible on the other, there is not one case of overlapping scores for pro and anti *S*s as determined by the Jones-Aneshansel scale. Every score for the antis is less than 1, and all but one of the pros' scores is greater than 1. Whether such a high degree of discrimination would be found using other attitude measures with learning tasks of relevant materials is a matter for further empirical study.

SUMMARY

Two closely related experiments were conducted to test the following hypotheses: (*a*) prosegregation *S*s learn plausible prosegregation and implausible antisegregation statements with less difficulty than they learn plausible antisegregation and implausible prosegregation statements; (*b*) the reverse is true of the antisegregation *S*s; and (*c*) those with neutral or intermediate beliefs learn all types of statements equally well. Each hypothesis was confirmed, and the reliability of the findings was especially striking in the replication experiment where the amount learned was more carefully scored, and other precautions were taken against biasing the results.

While the results show that learning of a controversial statement cannot be predicted solely from the direction of the argument, they still support the functionalist's basic assumption that cognitive processes operate so as to promote the constancy of attitude and belief.

REFERENCES

1. ALPER, THELMA G., & KORCHIN, S. J. Memory for socially relevant material. *J. abnorm. soc. Psychol.*, 1952, **47,** 25–37.
2. BRUNER, J. S., & POSTMAN, L. Emotional selectivity in perception and reaction. *J. Pers.*, 1947, **16,** 69–77.
3. EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
4. EDWARDS, A. L. Political frames of reference as a factor influencing recognition. *J. abnorm. soc. Psychol.*, 1941, **36,** 34–50.
5. HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion, psychological studies of opinion change.* New Haven: Yale Univer. Press, 1953.
6. HOVLAND, C. I., & PRITZKER, H. A. Extent of opinion change as a function of amount of change advocated. *J. abnorm. soc. Psychol.*, 1957, **54,** 257–262.
7. JANIS, I. L., & KING, B. T. The influence of role-playing on opinion change. *J. abnorm. soc. Psychol.*, 1954, **49,** 211–218.
8. JONES, E. E., & ANESHANSEL, JANE. The learning and utilization of contravaluant material. *J. abnorm. soc. Psychol.*, 1956, **53,** 27–33.
9. LAMBERT, W. W., SOLOMON, R. L., & WATSON, P. D. Reinforcement and extinction as factors in size estimation. *J. exp. Psychol.*, 1949, **39,** 637–641.
10. LEVINE, J. M., & MURPHY, G. The learning and forgetting of controversial material. *J. abnorm. soc. Psychol.*, 1943, **38,** 507–515.
11. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
12. MCCLELLAND, D. C., & LIBERMAN, A. M. The effect of need for achievement on recognition of need-related words. *J. Pers.*, 1949, **18,** 236–251.
13. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56,** 244–251.
14. POSTMAN, L. The experimental analysis of motivational factors in perception. In J. S. Brown et al., *Current theory and research in motivation.* Lincoln: Univer. Nebraska Press, 1953.
15. POSTMAN, L., BRUNER, J. S., & MCGINNIES, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **43,** 142–154.
16. PRATT, C. C. The role of past experience in visual perception. *J. Psychol.*, 1950, **30,** 85–107.
17. SCHAFER, R., & MURPHY, G. The role of autism in visual figure-ground relationship. *J. exp. Psychol.*, 1943, **32,** 335–343.
18. TAFT, R. Selective recall and memory distortion of favorable and unfavorable material. *J. abnorm. soc. Psychol.*, 1954, **49,** 23–28.
19. VANDERPLAS, J. M., AND BLAKE, R. R. Selective sensitization in auditory perception. *J. Pers.*, 1949, **18,** 252–266.
20. WATSON, W. S., & HARTMAN, G. W. The rigidity of a basic attitudinal frame. *J. abnorm. soc. Psychol.*, 1939, **34,** 314–335.

Received August 16, 1957.

VICARIOUS EXPERIENCE AND PARTIAL REINFORCEMENT[1]

DONALD J. LEWIS[2] AND CARL P. DUNCAN

Northwestern University

PSYCHOLOGISTS and others have used a whole cluster of concepts when referring to one person, in some fashion, "taking the part of another." Among these concepts are "imitation," "copying," "identification," "the significant other," "the normative generalized other," "role," "role playing," etc. These terms refer to different manners and different degrees of vicarious functioning, going, along one dimension, from a rather mechanical following of another's movements to the "internalization" of another's value systems and emotions so that one person can almost live for the other.

The theoretical history of various forms of vicarious functioning is a long one, but the history of experimentation on the problem is very brief. One of the first theoretical and experimental approaches to the problem from a behavioral science point of view was that of Miller and Dollard (7), who distinguished between "imitation" and "copying." Imitation was the matching of subjects' (*S*s') responses to those of a human model. Because the model's presence was a necessary condition for imitation, this kind of behavior was characterized as matched-dependent. "Copying" referred to *S*'s following the model's behavior, or his supposed behavior, partly on the basis of cues produced by the copier's own responses, even when the model was absent.

Osgood (9) and Lazowick (3) indicate that the basis of copying and identification (putting oneself in the place of another) is a mediational response (r_M) common to both model and *S*. A mediational response is behavior, frequently internal or unobserved, which intervenes between environmental stimuli and instrumental responses. A simple diagrammatic representation of a mediational response is indicated by the following: $S \rightarrow r_M \rightarrow s_M \rightarrow R$. "S" represents the external, environmental stimuli, and "r_M" represents an internal, often autonomic, response to S. Because r_M is a response, it therefore has the properties of all responses, one of which is the emission of stimuli. The stimuli emitted by r_M are symbolized as "s_M" and these lead directly to and are an important component of the total stimulus situation, which also includes S, evoking the final instrumental response R.

The concept of mediational response has been highly developed in the area of learning by Hull (1), Spence (11), and others. In Spence's theory, the theoretical concept K is regarded as representing the motivational properties of the r_M-s_M mechanism and is defined in terms of specific empirical variables, viz., the magnitude of reward and the delay of reward. The mediational response (fractional anticipatory goal response) has served S-R theory in some of the ways that "expectancy" has served Tolman and his followers. Spence, for example, has shown how latent learning can be derived with the help of r_M-s_M. Osgood (8) has indicated that *meaning* probably has its source in a mediating response, and that when an environmental S means the same thing to two different persons it is because the S evokes the same or similar r_Ms. Amsel (12) has used r_M to interpret some of the data of partial reinforcement, and the concept has also been used to predict new data as illustrated by the study of Kendler, Karasik, and Schrier (2).

The present study represents an attempt to determine some of the conditions under which one can benefit from the experience of others and to relate these conditions to S-R theory, using a mediational response mechanism.

Previous research (4) has shown that partial reinforcement effects can be obtained from humans *S*s in a lever-pulling ("slot-machine") situation; those *S*s who win all of the time (continuously) during an acquisition series quit more quickly, when given the opportunity, than do *S*s who have won only part of the time and irregularly (partially). This study attempts to determine if these partial reinforcement effects can be obtained from *S*s whose acquisition experience in the situation is purely vicarious, resulting either from watch-

---

[1] This study was supported by a grant to the authors from the National Science Foundation.

[2] Now at Louisiana State University, Baton Rouge, Louisiana.

ing someone else play or from simply hearing about the experiences of someone else who has played. The theoretical interpretation of what happens during this vicarious experience follows most closely that of Osgood (9) and Amsel (12).

Consider that *E*, acting as a model, plays the slot machine and that *S* watches him play. The external stimulus situation S is therefore very similar for both. Assume further that both *S* and *E* like to win and dislike to lose. In addition, they both know how to put coins in slots and pull levers, and these simple instrumental responses, from much past experience, are under the command of verbal stimuli. That is to say, all adult *S*s can follow, completely without experience in the present specific situation, the verbal instructions, "Put a coin in the slot and pull the lever."

All *S*s have in common, then: (*a*) the same external situation; (*b*) the same highly developed skill at the instrumental response of lever pulling; (*c*) the same verbal command over this response, and (*d*) the same emotional responses, positive for winning and negative for losing. The external situation is S, the instrumental response is R, and both the emotional responses and the verbal commands are r_Ms.

With this much in common for all *S*s, it is necessary to note the differences that develop as a function of the experimental variables. These differences will be most clear if we consider first only the person who actually plays the machine, *E* in this case. When *E*, under a continuous reinforcement schedule, pulls the lever and wins, a pleasurable (at least distinctive) emotional response results. This mediating response, resulting from reward, will be symbolized as r_R. The r_R yields stimuli—s_R—which, among others, lead to the instrumental response of the next lever pull and thus s_R becomes attached to lever pulling.[3] When the extinction series starts and lever-pulling responses are no longer rewarded, a different r_M occurs, one that has a displeasurable or frustration component—r_F. The occurrence of r_F and its associated stimuli s_F results in, among other things, a considerable change in stimuli during extinction from those that had been present during acquisition, and extinction proceeds rapidly.

For the partially reinforced *E*s, the situation is somewhat different. Both winning and losing occur during acquisition and thus both r_R-s_R and r_F-s_F occur during acquisition. Since both s_R and s_F are followed by lever pulling, the instrumental response becomes attached to both the pleasurable and unpleasurable responses, and quitting is delayed when the extinction series starts.[4]

The theoretical interpretation is thus clear enough for *E* who actually plays the machine; it is not quite so clear for the *S*s who simply watch. Although it is not difficult to see that r_R and r_F occur in these *S*s, it is the connection of r_R-s_R and r_F-s_F to the lever-pulling response that presents a problem because *S* never actually pulls the lever. The solution lies in the fact that the instrumental lever-pulling response is strong in the response repertoire of both *S* and *E*, and this instrumental response is led to by such subvocal responses (and their associated stimuli) as, "Good! A win. Now it is time for the next trial. Pull the lever." These verbal mediational responses have been associated with lever pulling through much past experience.

If we assume that the external stimulus situation, as a result of the events of the previous trials, evokes the appropriate r_Ms (r_R or r_F) in both *S* and *E* and that the r_M leads to, and therefore gets associated with, the subvocalizing about the next trial, and if the subvocalizing is already hooked to the instrumental response, then a definite chain is established between stimulus and instrumental response in the *S*s who simply watch as well as in the *E*s who actually play. A diagram of the chain would look like this:

$$S \rightarrow r_M \rightarrow s_M \rightarrow \text{verbalization} \rightarrow R.$$

The verbalization → R part of the chain has been learned previous to the experiment by *S*, but the $S \rightarrow r_M$ link, and the $s_M \rightarrow$ verbalization link are learned here during the acquisition series. The series of events occurring during the experiment can be said to have the same meaning for both *S* and *E* because the same r_Ms occur.

---

[3] Assuming a continuity theory of reinforcement.

[4] There are undoubtedly other reasons for the superior resistance to extinction following partial reinforcement. One is the differential change in orienting or observing responses (13) that take place as a result of these two kinds of reinforcement.

The preceding analysis holds for those *S*s who watch someone else play the machine. For those who simply hear a description of the play, the analysis is only slightly different. The external situation for them consists primarily of verbal stimuli; the description of what happens. These verbal stimuli, it is assumed, will also be able to evoke the relevant r_Ms. Thus a diagram of the chain of events is this: verbalization $\rightarrow r_M \rightarrow s_M \rightarrow$ verbalization $\rightarrow$ R.

To bring data to bear on these theoretical notions and on an important social psychological problem, a ten-group experiment was performed. Two of the groups were traditional partial reinforcement groups with 25% and 100% being the percentages of reward involved. Two groups (one with 25% and one with 100% reward) watched *E* play the acquisition series and won on the plays that *E* did. It was believed that actually winning would increase *S*s' degree of participation in the situation. Two more groups (25% and 100%) watched *E* play the acquisition series but did not participate in the winning. Two groups had a hypothetical acquisition series explained to them and won when the *E* hypothetically won. A final two groups had a hypothetical acquisition series explained to them, but the *S*s did not participate in the winning. After the acquisition series was actually or vicariously completed, *S* took over the machine and was told he could play as long as he wanted. Two response measures were taken. The primary one, and the one to which the theoretical treatment refers, is the number of plays to quitting. Also a measure of *S*'s expectation on each trial was taken.

METHOD

Apparatus

The apparatus has been described in detail elsewhere (5). It is a modified slot machine. The player puts a metal disk in a slot at the top and then pulls a lever. A whirring sound ensues, lights flash, and a disk either comes out of a payoff tray or does not. The player has access to a large supply of these disks and can play as long as he desires. There is a counter inside the machine which cumulates the total number of plays. The machine pays off on particular plays which are determined in advance by the *E*; payoffs do not depend on anything *S* does.

Design

Five pairs of groups were used, with one group in each pair having a 25% reward acquisition series and the other a 100% reward acquisition series. For one pair, the Control, *S* played the acquisition series himself, and, of course, went right on to extinction. For another pair, Watch and Win, *S* watched *E* play the acquisition series. Whenever *E* received a payoff disk, *S* was given one too. A third pair, Watch Only, differed from the Watch and Win pair only in that *S*s never received a payoff. A fourth pair, Explain and Win, was told orally about a hypothetical person who had just previously played the machine. The *S* was given a play-by-play description of the hypothetical person's acquisition series. When the hypothetical *E* won, *S* was given a disk too. The final pair was the Explain Only. The same explanation was given to this group as to the Explain and Win, but now the *S*s were not given disks at any time. The acquisition series consisted of the first eight trials, with the 100% *S*s of each pair being paid on each of the eight trials and the 25% *S*s being paid on only Trials 2 and 8. After the eighth trial, *S* played the machine until he said that he wanted to quit, but he never won again. He was told that all disks that he had in his possession could be turned in when he was finished for 10¢ each. This exchange, of course, actually took place. No *S* in any group knew that the machine would never pay off (extinction) after the eighth trial, no matter whether the first eight trials had been played by *S*, *E*, or some hypothetical third person.

The *S*s were also asked to state on each play, just before they pulled the lever, whether they expected to win or not. Secured to the front of the machine was a line-rating scale with the numbers 1 through 6 placed at equal intervals on the line. Above these numbers was the title, "Expectation of winning or not winning." Corresponding to the numerical scale points from 1 through 6 were the descriptions: "fairly sure of not winning," "mildly sure of not winning," "slightly sure of not winning," "slightly sure of winning," "mildly sure of winning," and "fairly sure of winning." On each trial, before he pulled the lever, *S* called out the numerical value corresponding to his level of expectancy. The *E* sat to one side of the room and recorded the expectancy stated on each trial by *S*.

Subjects

The *S*s were drawn from introductory psychology courses at Northwestern University and were assigned to the various experimental groups as they appeared in the laboratory. Twenty *S*s were assigned to each of the 10 experimental conditions, making a total of 200.

RESULTS

Mean Log Plays to Extinction

The total number of plays for each *S* was converted into common logarithms and the mean for each group determined. A simple t was computed between the 100% and 25% mean log plays for the Control, and the usual partial reinforcement effect was found (t = 2.36 for 38 *df*). Then a succession of four 2 × 2 factorial analyses of variance was computed, comparing each of the experimental groups against the control. One such comparison, for example, had 25% and 100% for the columns, and the Control condition against the Watch and Win condition for the rows. Another comparison had 25% and 100% for the columns, and the Control condition against Explain and Win for the rows.

The comparisons of Watch and Win against Control showed no difference. Vicarious experience, in other words, was as effective as the actual playing. The Watch and Win group actually played longer, but insignificantly so, than the Control. The percentage of reinforcement was significant ($F = 4.90$ for 1 and 76 *df*), with more plays to extinction for the 25% groups (see Fig. 1) and the interaction was not ($F = 2.54$ for 1 and 76 *df*). Thus the partial effect occurred in both the vicarious groups and the actually playing group, and not differentially.

The comparison of Explain and Win against Control also showed no significant difference between the actual and the vicarious experience groups. The percentage of reinforcement was significant at the .01 level ($F = 7.33$ for 1 and 76 *df*), and again there was no significant interaction ($F = 2.05$ for 1 and 76 *df*).

The comparisons of Watch Only against Control, and Explain Only against Control gave no significant main effects, but in both cases there was a significant interaction at the .05 level. This interaction is due to the effect that percentage of reward had on the control group but not on the experimental groups.

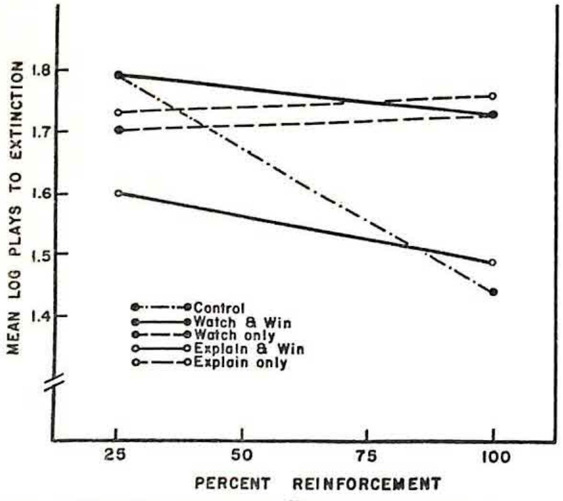

FIG. 1. THE RELATIONSHIP BETWEEN PERCENTAGE OF REINFORCEMENT AND MEAN LOG PLAYS TO EXTINCTION

As far as mean log plays to extinction are concerned, then, the vicarious groups did as well as the control groups overall. The partial reinforcement effect, however, appeared only in those situations where the *S*s actually won during the acquisition series.

Expectancies During Extinction

Because *S*s played different numbers of extinction plays, the extinction trials were Vincentized into ten blocks of trials for each *S*. The expectancy score for each *S* on the ten extinction blocks consists of the sum of his expectancies scores over that block of trials divided by the number of trials in the block. Therefore, the score for any block can be any value from one to six, the same values as in the original scale. Any score over three represents an expectancy that the machine will pay off on the next trial; a score below three represents an expectancy that the machine will not pay off on the next trial.

A three-point "trial-by-trial" trend analysis was performed on the data with the mean of the first two Vincentized tenths, the mean of the fifth and sixth tenths, and the mean of the last two tenths serving as the three points for the trend. Again a succession of 2 × 2 analyses of variance compared each of the experimental groups against the control for the trend. There were no significant main effects for any of the comparisons, but several interactive trend effects emerged as significant.

For the comparisons concerning Watch and Win vs. Control, and Explain and Win vs. Control, the trials by percentage interaction were significant at the .05 level. A graph of these data (not presented here) shows that the expectancy curve after 100% reinforcement falls off faster and ends up lower than does the curve after 50% reinforcement. For the comparisons with respect to Watch Only vs. Control, and Explain Only vs. Control, percentage of reinforcement was not an effective variable on trend of expectancies. In order for there to be a differential change of expectancies during extinction it was necessary for the *S*s actually to win during the acquisition series.

For all four of the 2 × 2 comparisons there was a significant triple interaction, trials by percentage by experimental condition. In each case, this is interpreted to mean that percentage of reinforcement had a greater effect upon the control group expectancies than upon the experimental group expectancies.

DISCUSSION

The evidence indicates that the experimental variables had similar effects upon both response measures—mean log plays to extinction, and extinction expectancies; there were no over-all differences between any of the vicarious groups and the controls, e.g., vicarious and control *S*s showed the same over-all number of plays to extinction. The partial reinforcement effect—a greater number of plays to extinction for the 25% *S*s over the 100% *S*s—occurred, however, only in those vicarious groups which had participated in the winning. Previous research (5, 6) has revealed a characteristic pattern of extinction expectancies as a function of partial reinforcement; that pattern was found here, but only for the *S*s who participated in the winnings. It seems clear that if *S* did win in the situation, even though he never pulled a lever during acquisition himself, the partial effect showed clearly during extinction. It made no difference under these conditions whether *S* watched someone else play or was merely told what someone else had done while playing.

The theory that prompted this study predicted that the vicarious acquisition of the partial effect would occur, and the theory is to this extent confirmed. It was, however, originally expected that the Watch Only and the Explain Only groups would also show the partial effect. That they did not can possibly be attributed to the subtlety of the partial effect. It is an effect which is not under *S*'s verbal control, in that most *S*s do not know about the effects of partial reinforcement. One of the writers has asked several large classes to predict which of two persons would persist longer in the face of failure in a gambling situation; the one who won all the time or the one who won part of the time and irregularly. The classes split about half-and-half in their predictions. Apparently common sense does not make most people aware of what has become one of the firmest laws of psychology. It seems clear, nevertheless, that a nonverbalized experience can be imparted vicariously.

Another reason why the nonwinning groups did not show the partial effect may lie in the lower degree of participation which these groups experienced. The theory ties the partial effect closely to a mediating response, and it was guessed that this mediating response is emotional (motivational). It may well be the case that when *S* actually wins he is more of a participant in the situation and the emotional responses from winning and not winning are more likely to occur. And they must occur, it is assumed here, in order for them to get associated with the instrumental response.

Although this experiment was stimulated by a mediational response theory, and it is believed that the data are not inconsistent with such a theory, it cannot be doubted, nevertheless, that other theoretical points of view may also be able to handle the data, perhaps as well. Further, there are problems presented by these data that mediational response theory offers no clear-cut interpretation for. Other experiments are in progress in the attempt to solve some of them.

SUMMARY

Ten groups (five pairs) of subjects worked a modified slot machine, half under a 100% reinforcement schedule and half under a 25% reinforcement schedule. The first eight trials constituted the acquisition series, and the remainder of the trials were extinction. One pair of groups watched *E* play the acquisition, winning when *E* did. Another pair watched *E* play the acquisition series but did not participate in the winning. A third pair was told trial-by-trial, about a hypothetical person's acquisition series, winning when the hypothetical person won. A fourth pair was also told about the hypothetical person's play but did not share in the winning. The *S*s of a fifth and final pair played the acquisition series themselves. All *S*s then played the extinction series, being told that they could play as long as they wished.

The results showed that vicarious learning did take place as long as *S*, himself, actually participated in the winning during acquisition. The explanation of the results was offered in terms of a mediational response.

REFERENCES

1. HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
2. KENDLER, H. H., KARASIK, A. D., & SCHRIER, A. M. Studies of the effect of change of drive: III. Amounts of switching produced by shifting drive from thirst to hunger and from hunger to thirst. *J. exp. Psychol.*, 1954, **47**, 179–182.

3. LAZOWICK, L. M. On the nature of identification. *J. abnorm. soc. Psychol.*, 1955, **51,** 175–183.
4. LEWIS, D. J., & DUNCAN, C. P. Effect of different percentages of money reward on extinction of a lever-pulling response. *J. exp. Psychol.*, 1956, **52,** 23–27.
5. LEWIS, D. J., & DUNCAN, C. P. Expectation and resistance to extinction of a lever-pulling response as a function of percentage of reinforcement and amount of reward. *J. exp. Psychol.*, 1957, **54,** 115–120.
6. LEWIS, D. J., & Duncan, C. P. Expectation and resistance to extinction of a lever-pulling response as a function of percentage of reinforcement and number of acquisition trials. *J. exp. Psychol.*, 1958, **55,** 121–128.
7. MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.
8. OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49,** 197–237.
9. OSGOOD, C. E. Behavior theory and the social sciences. *Behav. Sci.* 1956, **1,** 167–185.
10. SPENCE, K. W., BERGMANN, G., & LIPPITT, R. A study of simple learning under irrelevant motivational-reward conditions. *J. exp. Psychol.* 1950, **40,** 539–551.
11. SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.
12. WILSON, W., WEISS, E. J., & AMSEL, A. Two tests of the Sheffield hypothesis concerning resistance to extinction, partial reinforcement, and distribution of practice. *J. exp. Psychol.*, 1955, **50,** 51–65.
13. WYCKOFF, L. B., JR. The role of observing responses in discrimination learning. *Psychol. Rev.*, 1952, **59,** 431–442.

Received August 23, 1957.

THE EXPRESSION AND INHIBITION OF HOSTILITY IN IMAGINATIVE AND OVERT BEHAVIOR[1]

JACK E. HOKANSON

University of Wisconsin

AND JESSE E. GORDON

Montana State University

SEVERAL recent studies (1, 2, 4) have used extensions of Miller's approach-avoidance paradigm to deal with behavior on projective thematic materials. In general, the findings of these studies reveal that with increased cue relevance and/or heightened motivation, there is a diminution in the projected expression of the particular motives being studied. The explanation offered by these authors is that the anxiety aroused by the heightened stimulation produces an inhibitory effect on the direct expression of the motive in the subject's (*S*'s) fantasy productions. The motives studied in this framework were hunger (2), sex (1), and threat avoidance (4).

The present study attempts to investigate the expression of hostile motivation in both imaginative and overt behavior, using the goal gradient formulation as a source of hypotheses. Three sets of variables pertaining to the expression of hostility in imaginative behavior were studied in a factorial design; in a parallel arrangement, these same variables were studied with regard to the overt expression of hostility. The variables studied were: (*a*) high and low provocation to anger (the Arousal variable), (*b*) high and low cue (a TAT card that normally elicits hostile fantasy vs. one that does not), (*c*) individual differences in the tendency to admit to or express hostility in a paper-and-pencil test situation (Expressor variable).

There are two goal-gradient models on which predictions might be based; these models lead to opposite predictions. The first alternative (Fig. 1) depicts Low Expressors as being at Point A. As the cue or motivational conditions for hostility are increased to A′, this group of *S*s should manifest a net increase in the expression of hostility; they have not yet reached a point where the inhibitory gradient is operative. This model depicts High Expressors as being at Point B. As the cue or motivational conditions for hostility are increased to Point B′, this group should show a net decrease in the expression of hostility, since the inhibitory gradient is assumed to be rising sharply in this area. Thus one would predict from this model that under conditions of high cue and arousal, people who do not ordinarily readily admit to hostility should increase their hostile expression, and those who readily admit to hostility should mainfest a decrease in hostile expression. This model assumes that people who do not readily admit to hostile feelings have less hostile motivation (are further from the goal) than those who report hostile feelings.

An alternative, though somewhat less parsimonious, model leads to opposite predictions (Fig. 2). In this formulation, the High and Low Expressors are assumed to have equal motive strengths for hostility but to differ in the degree to which they inhibit expression of this hostility. Thus both groups are considered to be at Point A, but the inhibitory gradients of the two groups differ. The inhibitory gradient of the Low Expressors is already operative, allowing for their tendency to not admit hostility readily on a paper-and-pencil test. As the cue or motivational conditions are increased to Point B, these Low Expressors should inhibit even more. The inhibitory gradient of the High Expressors, on the other hand, is not yet operative at Point A. Thus these *S*s readily admit to hostility on the paper-and-pencil test. As cue and arousal conditions are heightened, these *S*s are moved to a point where their inhibitory gradient is still low. Thus these *S*s should manifest an increase in hostile expression as they are moved to the left on the abscissa.

Exactly opposite predictions therefore follow from these models, and two-tailed tests are called for. While requiring the additional assumption that High and Low Expressors

---

[1] This research was conducted at the University of Wisconsin.

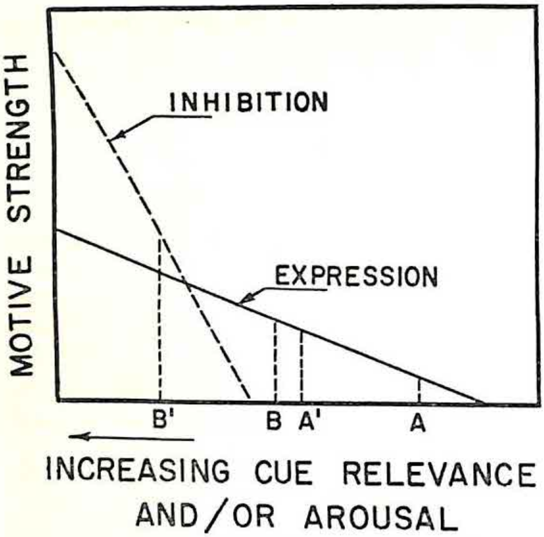

FIG. 1. MODEL BASED ON BETWEEN-GROUP DIFFERENCES IN MOTIVE STRENGTH

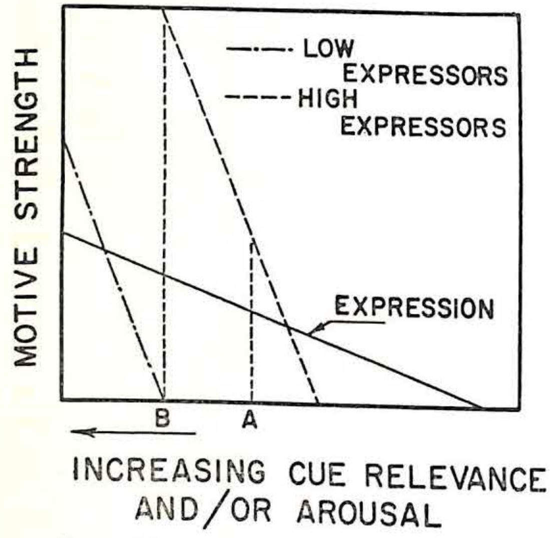

FIG. 2. MODEL BASED ON BETWEEN-GROUP DIFFERENCES IN AVOIDANCE GRADIENTS

differ in their inhibitory gradients, the second model seems to agree with the clinical observations that people who are very docile and nonaggressive are often inhibiting strong underlying hostility. Clinical experience therefore suggests that specific hypotheses be drawn from the second model. The hypotheses to be tested are:

1. People classified as Low Expressors manifest a net decrease in hostile expression, in both imaginative and overt behavior, as the cues and arousal conditions for hostility are increased.

2. People classified as High Expressors manifest a net increase in hostile expression, both in imaginative and overt behavior, as the cues and arousal conditions for hostility are increased.

The general outline of the experimental design is as follows: The extreme high and low scorers on Segal's Manifest Hostility Scale (5) were classified as High and Low Expressors. Half of each group was placed in a situation designed to arouse strong hostility and the other half placed in a low arousal situation. Hostility was measured from TAT protocols administered subsequent to arousal. Cards that usually elicit considerable hostile fantasy (High Cue) and those that do not (Low Cue) were administered to each *S*. Hostility was also measured in a behavioral situation in which *S* was given the option of administering an electric shock to *E*. Here again a device was used to provide high and low cue stimulation.

Thus a three-way classification analysis of variance was applied to both the fantasy and the overt measures of hostility. Each *S* appeared in both High and Low Cue conditions; a repeated measures analysis was used to take these correlated measures into account. The predictions made should be reflected in Expressor × Arousal and Expressor × Cue interactions.

METHOD

Forty male white undergraduate students at the University of Wisconsin taking an introductory course in psychology served as *S*s in this experiment. Their ages ranged from 18 to 26. *S*s were selected on the basis of their scores on the Segal Manifest Hostility Scale. The 20 *S*s with low scores (below 8) formed the Low Expressor group, and the 20 *S*s with high scores (above 18) comprised the High Expressor group. Segal's Manifest Hostility Scale, an objective test based on MMPI items, is described in detail elsewhere (5).

Each *S* was seen individually for approximately 40 minutes. Half of the *S*s in each group were put into an hostility-arousing situation (High Arousal), and the other half put into a neutral situation (Low Arousal). In the High Arousal condition, *S*s were asked to listen to a case study of a sadistic juvenile delinquent. Care was taken to minimize the possibility of sympathetic reactions to the delinquent, and to avoid the implication that the delinquent was mentally ill. From the *S*'s point of view, the case study portrayed a person who "simply enjoyed hurting others." This procedure was designed to produce feelings of punitiveness and hostility toward the delinquent. A case history of a juvenile delinquent who was portrayed as an innocent victim of circumstance and environmental forces was read to the Low Arousal group. The presentation was designed to elicit sympathy for the delinquent.

Immediately after the case presentation, each *S* was asked to answer three questions designed to check whether the experimental manipulations had their desired effects. These questions required *S*s to predict the punishment received by the delinquent in the case study, to indicate what punishment *S* felt to be appropriate, and to indicate how sympathetic *S* felt toward the delinquent.

Following the above procedures, *S*s in both conditions were told that the research involving the delinquent was concluded but that *E* was interested in some other things. *S* was asked to compose imaginative stories to two TAT cards. Each *S* was presented with Cards 2 and 8 BM and asked to write a story for each one, with a five-minute time limit for each card. The order of presentation of the cards was counterbalanced with respect to the Expressor groups and the Arousal conditions. Cards 2 and 8 BM were specially selected for their difference in relevance to hostility. A normative study by Eron (3) reveals that Card 8 BM normally elicits a great deal of hostile and destructive fantasy (51.31% of responses), whereas Card 2 does not usually elicit such fantasy (6.9% of responses). Thus Card 8 BM served as the High Cue card, and 2 as Low Cue.

Scoring procedures for the TAT stories were set up prior to the experiment. Cards 2 and 8 BM had been administered to a discussion section of an intermediate psychology class, and 21 themes related to hostile expression extracted from the protocols obtained. These themes were ranked by graduate students in psychology for intensity of hostile expression. The themes were then arranged according to mean ranking and assigned to seven categories of three themes each. These categories formed the criteria for an eight-point rating scale of hostile expression on the TAT, on which a score of zero signifies a story completely devoid of hostility, and a score of seven signifies intense hostility.

A group of protocols was scored independently by *E* and two other persons. Interscorer reliability with *E* was $r = .79$ and $r = .80$. It was therefore felt that the scale scoring was reasonably reliable, and *E* proceeded to score the TAT protocols of the experimental *S*s.

The method for measuring overt hostility was one in which *S* had the option of administering electric shocks to *E*. Following the TAT administration, *E* introduced this situation with the following instructions:

> The third part of this session is again something quite different. As you are probably well aware, one of the biggest sources of error in the interpretation of psychological tests is the interpreter himself. For instance, if I were to interpret your TAT stories, such factors as being in a bad mood or feeling physically ill would certainly affect my interpretation. This is the problem I am attempting to investigate in this part of the session. Now, I am going to read over your TAT stories and make a quick interpretation to myself. Then I am going to read out loud some items from a type of personality questionnaire and try to predict how I think you would answer each item. If my prediction is wrong I want you to give me an electric shock through this apparatus. If I am correct in my prediction, then simply don't do anything and I will go on to the next item. In this way I will feel physical discomfort when my predictions are wrong; I would like to see how this affects my subsequent predictions.

The questionnaire was one developed to measure attitudes toward dating at the University of Wisconsin. The items are rather straightforward, innocuous statements which appear to keep personal involvement at a minimum. Thus to either agree or disagree with a particular item should make relatively little difference to the *S*. The nature of the items were such that regardless of *S*'s feelings about a particular item, the decision to shock or not shock was mostly determined by his readiness to express hostility overtly. It was reasoned, therefore, that the number of times *S* disagreed with *E*'s prediction (i.e., number of times *S* shocked *E*) could be used as a measure of overt hostile expression.

The questionnaire consisted of 18 items on which at least 80% of the standardization groups of University of Wisconsin male students gave the same response. Each item could be answered by Agree, Disagree, or Neutral. Items in which the agreed upon response was "neutral" were not used. Each prediction was made in a direction different from that which the normative population indicated to be the prevailing attitude. In this way, each *S* was faced with 18 predictions about himself with which he most likely disagreed. Thus he was given 18 cues to shock *E*. The items were divided into High and Low Cue items in order to parallel the design used in the fantasy measures. A High Cue item was one in which *E* predicted a response exactly opposed to the prevailing attitude expressed by the normative population, and by implication, opposed to *S*'s own attitude. A Low Cue stimulus was a prediction of "neutral," when the prevailing attitude of the norm group was either Agree or Disagree.

Each *S* received 9 High Cue and 9 Low Cue "predictions." The sequence and cue value of the items were balanced systematically between the High and Low Expressors and between the High and Low Arousal groups.

RESULTS

Tables 1, 2, and 3 reveal that the reading of the case studies was successful in arousing hostility and punitiveness. Table 1 reveals a highly significant difference between the predicted punishment for the delinquent in the High Arousal case and for the delinquent in the Low Arousal case. The mean predicted punishment score in the High Arousal condition was 3.2 (a jail sentence of approximately 16 years) while the predicted punishment in the Low Arousal was 1.4 (approximately 7 years). There was no significant difference between Low and High Expressors in their estimates of punishment, nor was there a significant Expressor × Arousal interaction. This finding suggests that the amount of arousal was similar for both High and Low Expressor groups.

Table 2 reveals that the *S*s desired to punish the delinquent in the High Arousal case significantly more severely (approximately 27 years) than the delinquent in the Low Arousal case (approximately 3½ years in prison). There were no significant differences between Low and High Expressors in their desires to have the delinquents punished, nor was there a significant Expressor × Arousal interaction.

Table 3 indicates that *S*s expressed significantly less sympathy toward the delinquent in the High Arousal case. On an eight-point rating scale of sympathy on which a high score reflects low sympathy, the delinquent in the High Arousal case received a mean of 5.15, while the delinquent in the Low Arousal case received a mean score of 2.7. Again there were no significant differences between High and Low Expressors, and the Expressor × Arousal interaction was again nonsignificant.

These consistent results indicate that the delinquent in the High Arousal group did indeed arouse considerably more punitiveness and hostility among the *S*s than he did in the Low Arousal group.

Table 4 summarizes the analysis of variance of the hostility scores obtained on the TAT. The significant Expressor × Arousal interaction is opposite in direction to that predicted in our hypotheses. These results indicate that in a low arousal situation, the High Expressors' fantasy productions contain somewhat more hostile content than those of Low Expressors. However, when arousal is increased, the relationship is reversed; the Low Expressors' fantasy increases in hostile content while the High Expressors' fantasy decreases in hostile content. In comparing the total scores of High and Low Expressors, as a main effect, no significant difference was found.

Table 4 also indicates that the cue value of the TAT cards had a significant effect on the amount of hostile fantasy elicited. The mean score on the Low Cue card was 1.95, and 2.87 on the High Cue card. The predicted interaction of Expressor × Cue was not significant.

Table 5 summarizes the analysis of variance

TABLE 1

PUNISHMENT OF DELINQUENT PREDICTED BY SUBJECTS

Source	*df*	Sum of Squares	Mean Square	*F*
Expressor	1	2.0	2.0	—
Arousal	1	32.4	32.4	14.33*
Expressor × Arousal	1	.6	.6	—
Within	36	81.36	2.26	
Total	39	116.36		

* $p < .01$, one-tailed test.

TABLE 2

PUNISHMENT OF DELINQUENT DESIRED BY SUBJECTS

Source	*df*	Sum of Squares	Mean Square	*F*
Expressor	1	6.8	6.8	2.00
Arousal	1	228.0	228.0	67.25*
Expressor × Arousal	1	5.29	5.29	1.56
Within	36	122.1	3.39	
Total	39	362.19		

* $p < .01$, one-tailed test.

TABLE 3

EXTENT OF SYMPATHETIC FEELINGS TOWARD DELINQUENT

Source	*df*	Sum of Squares	Mean Square	*F*
Expressor	1	.9	.9	—
Arousal	1	60.0	60.0	57.69*
Expressor × Arousal	1	1.6	1.6	1.53
Within	36	37.78	1.04	
Total	39	100.28		

* $p < .01$, one-tailed test.

TABLE 4

TAT HOSTILITY SCORES

Source	*df*	Sum of Squares	Mean Square	*F*
Expressor	1	.01	.01	–
Arousal	1	.31	.31	–
Expressor × Arousal	1	18.97	18.97	5.96*
Pooled *S*s groups	36	114.7	3.18	–
Cue	1	17.11	17.11	6.88**
Cue × Expressor	1	.62	.62	–
Cue × Arousal	1	.32	.32	–
Cue × Expressor × Arousal	1	.15	.15	–
Pooled *S*s × Cue/groups	36	71.2	1.97	
Total	79	223.39		

* $p = .05$, two-tailed test.

** $p = .02$, two-tailed test

TABLE 5

SHOCK SCORES

Source	*df*	Sum of Squares	Mean Square	*F*
Expressor	1	4.5	4.5	1.18
Arousal	1	1.0	1.0	–
Expressor × Arousal	1	3.6	3.6	–
Pooled *S*/groups	36	136.9	3.80	–
Cue	1	86.1	86.1	23.2*
Cue × Expressor	1	9.1	9.1	2.45
Cue × Arousal	1	5.5	5.5	1.48
Cue × Expressor × Arousal	1	9.2	9.2	2.47
Pooled *S*s × Cue/groups	36	133.9	3.71	
Total	79	390.5		

* $p < .01$, two-tailed test.

of the hostility scores obtained in the situation in which *S*s had the option of administering electric shocks to *E*. The only significant effect is that associated with the cue value of *E*'s predictions about *S*. When *E* gave a Low Cue (neutral) prediction, he was shocked on the average of 3.6 times by each *S*. When *E* gave a High Cue prediction (opposite in direction to the response *S* would probably have made), he was shocked on an average of 5.5 times by each *S*. The predicted interactions (Cue × Expressor and Arousal × Expressor) were not found to be significant. However, an interesting trend is in the same direction as the Arousal × Expressor interaction of the TAT scores. This interaction suggests that in going from Low to High Cue conditions, Low Expressors increase their frequency of shocking *E* more than do the High Expressors. The mean shock scores for Low Expressors increases from 3.5 to 6.3, while the High Expressors increase from 3.7 to 5.1.

DISCUSSION

The results depicted in Tables 4 and 5 are opposite in direction to the predictions made in Hypotheses 1 and 2. The Low Expressors increased in the amount of hostile fantasy expressed when moving from Low to High Arousal. The High Expressors decreased in the amount of hostile fantasy expressed in going from Low to High Arousal. A trend in the same direction, though not significant, was found on the shock scores in comparing Low and High Cue conditions. These results indicate that the assumptions underlying the placing of the expressive and inhibitory gradients for each group were in error. It was assumed that differences in scores on the Manifest Hostility scale reflected differences in the degree of inhibition of hostile impulses. This assumption implied that the motive strength behind hostile expression was similar for high and low scorers on the scale. Thus a model (Fig. 2) was selected that involved a single approach gradient, but different avoidance gradients for the high and low scorers on the scale. The results of this experiment suggest that the assumption of different avoidance gradients was not a valid one.

The alternative model proposed in Fig. 1 assumes that high and low scorers on the Manifest Hostility scale are operating on the same avoidance gradient, and that the differences in their scores are a function of differences in motive strength. In this model, the high scorers are placed higher on the approach gradient, and thus closer to the goal. The results of this experiment could be accounted for by this model. But the model presents another serious problem. With human *S*s, it is reasonable to expect that factors tending to inhibit hostile expression may operate throughout the entire range of cue and arousal conditions. That is, most people have probably received sufficient punishment for expressing hostility in the past that under any given set of cue and arousal conditions, some inhibitory potential is likely to be present. Figure 1, involving linear goal gradients, does not provide for this; inhibitory potential does not accrue until some time after the expressive gradient is above the zero-strength base line. A model is needed which leads to the same predictions, but also employs the concept of inhibitory potential more centrally.

In conflict studies involving the behavior of rats that have been shocked, a single avoidance motive is usually involved (escape from shock). Such a motive may be adequately represented by a straight line. However, in human *S*s operating under hostile motivation, it is reasonable to suppose that there are a number of avoidance motives called into play, such as fear of punishment from external sources, ambivalent feelings toward the object of aggression, fear of loss of esteem from others, fear of violation of the self concept, etc. Each of these avoidance motives may be represented by a separate inhibitory gradient, the slopes of which may vary. When summated into a combined inhibitory potential, the resultant gradient would be curvilinear rather than rectilinear.

In such a model the resultant inhibitory gradient may extend over the entire range of arousal conditions. The model takes into consideration the criticism cited above that a rectilinear gradient does not describe the inhibitory potentials that probably exist over the entire range of approach motive strengths in humans. Further, the results of this experiment are consistent with hypotheses which might be drawn from this model. Figure 3 presents this curvilinear model applied to the conditions of this experiment.

The Low Expressors are assumed to be at Point A on the basis of their scores on the Manifest Hostility scale. Those Low Expressors who are in a High Arousal situation move to Point A′, resulting in a net increase in hostile expression. High Expressors are considered to be at Point B on the basis of their scores on the Manifest Hostility scale. When placed in a High Arousal situation, these *S*s move the same distance to the left as do the Low Expressors (there were no differences between High and Low Expressors in the extent of arousal by the case studies). They therefore move to Point B′, which results in a net decrease in hostile expression.

Although the results of this experiment are consistent with either a model involving a rectilinear avoidance gradient or one involving multiple avoidance gradients resulting in a curved gradient, we feel that this latter model is more applicable to human conflict situations. This model also does without the seemingly arbitrary assumption of Miller and Dollard that the avoidance gradient must be steeper than the approach gradient. However, it does introduce other assumptions concerning the additivity of avoidance motive strengths, which must be validated empirically. Research on animals operating under multiple avoidance drives should shed light on the validity of the hypothesized curved resultant avoidance gradient.

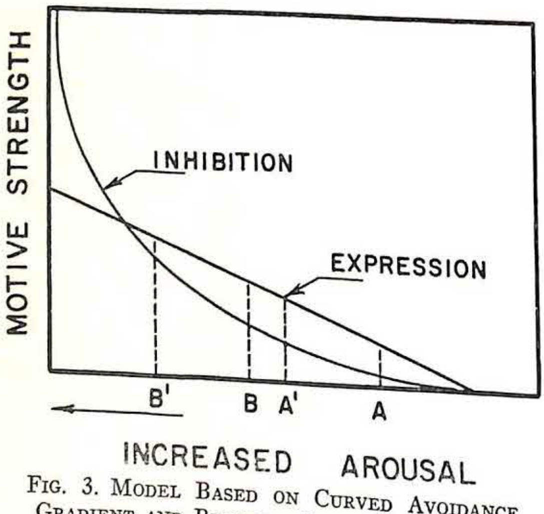

FIG. 3. MODEL BASED ON CURVED AVOIDANCE GRADIENT AND BETWEEN GROUP DIFFERENCES IN MOTIVE STRENGTH

SUMMARY

Half of a group of 40 white male college students, representing the extreme scores on the Segal Manifest Hostility Scale, were placed in a situation designed to arouse strong hostility, and half in a low arousal situation. *S*s were then permitted to express hostility in fantasy (using TAT pictures selected for differences in cue properties relevant to hostility) and in overt behavior, in a situation in which *S*s could actually hurt another person. As in the fantasy variable, the cue properties of the overt behavioral situation were varied.

Two alternative theoretical models based on Miller and Dollard's goal gradient formulation were proposed. Hypotheses derived from one of these models, which assumes differences in avoidance drive between high and low scorers on the Manifest Hostility scale, were as follows:

1. People classified as Low Expressors manifest a net decrease in hostile expression, both in fantasy and overtly, as the cues and arousal conditions for hostility are increased.

2. People who are classified as High Expressors manifest a net increase in hostile expression, both in imaginative and overt behavior, as the cues and arousal conditions for hostility are increased.

A significant Expressor × Arousal interaction in fantasy scores was in a direction opposite to that predicted, indicating that Low Expressors increased in hostile expression as arousal conditions increased, while High Expressors decreased in hostility under the same circumstances. No significant interactions involving differences in cue value were found, nor were there any significant interactions in the behavioral data.

The results of this experiment were consistent with a goal gradient model in which High and Low Expressors were assumed to differ in strength of approach motivation (proximity to the goal). However, criticisms of this model were detailed, and a substitute model, involving multiple avoidance gradients which summate to produce a curvilinear gradient, was proposed.

REFERENCES

1. CLARK, R. A. The effects of sexual motivation on phantasy. In D. C. McClelland (Ed.), *Studies in motivation*. New York: Appleton-Century-Crofts, 1955. Pp. 44–57.
2. EPSTEIN, S., & SMITH, R. Thematic apperception as a measure of the hunger drive. *J. proj. Tech.*, 1956, **20**, 372.
3. ERON, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64**, No. 9.
4. SCOTT, W. A. The avoidance of threatening material in imaginative behavior. *J. abnorm. soc. Psychol.*, 1956, **52**, 338–346.
5. SEGAL, S. M. The relationship of hostility to authoritarianism. *J. abnorm. soc. Psychol.*, 1956, **52**, 368–372.

Received August 17, 1957.

SIZE CONSTANCY AS A FUNCTION OF PERSONAL ADJUSTMENT AND DISPOSITION[1]

NOËL JENKIN

Training School at Vineland, New Jersey

MOST studies of size constancy, in relation to personality variables may be grouped into two categories. One group (2, 8, 9, 10) has claimed to show a relationship between the tendency to size constancy and some kind of personality disorder. The other group (11, 12, 15) has reported some relationship between constancy and the personality disposition of introversion-extraversion. Differing theoretical positions have instigated most of the inquiries, with the result that in each instance the independent variable has been a trait or type of diagnosis within only one or other of these two groups. There is clearly a need for further work in which variables representative both of adjustment and disposition are simultaneously studied in relation to size-distance judgment.

The direction of the relationship with a given personality variable is a matter which the literature shows to be ambiguous. Sanders and Pacht (10) found that the degree of size constancy increased concomitantly with gross degree of personality disturbance, from normality through neurosis to psychosis. Raush (9) obtained results consistent with this finding, but Lovinger (8) found that schizophrenics in poor contact with reality showed lower than normal size constancy under minimal cue conditions. Crookes (2) has also recently shown that neurotics and schizophrenics respond with a lower degree of size constancy than normals. Investigations employing the variable of introversion are somewhat more consistent, Weber (15) and Singer (11) finding that "extraverts" manifest a higher degree of size constancy than do "introverts," these results agreeing with the rather slight evidence of a similar nature earlier presented by Thouless (12).

It must be noted that there is only a broad similarity among the independent variables of each group of studies. Indeed, it may be questioned in some instances whether the variable has been correctly named. For example, Singer's (11) "thinking introverts" would be identified by Eysenck (3, 4) in terms of his "neuroticism" factor. If such a re-identification is valid, Singer's, and Sanders' and Pacht's results are in conflict, and this fact would add one more note of dissonance to an already inharmonious situation.

A possible explanation for at least some of the discrepancies lies in the nature of the instructions given to the subject (*S*). Sanders and Pacht, for example, used "objective" instructions, i.e., *S*s were ". . . to select the standard . . . which was equal in size to the stimulus square." Singer (11) used "Look" instructions, requiring that *S* ignore field cues and avoid the use of knowledge.

The present experiment was designed with the aim of bringing partial order into this area by varying the types of instructions given *S*, distinguishing groups of poorly adjusted and normal *S*s, providing the opportunity for classifying them as introverts and extraverts, and controlling for factors of age, intelligence, and sex (12).

METHOD

Apparatus

The apparatus for obtaining measures of size-distance judgment consisted of a movable white equilateral triangle which was raised and lowered through an aperture against a black mounting board. From *S*'s viewing position, the triangle when fully lowered was obscured below the aperture. A pulley and counterweight system enabled it to be exposed in graduated degree above the aperture, as indicated by a scale at the rear. *E* sat behind the board, thus screened from *S*'s view, and manipulated the size of the triangle, which at its largest setting measured 27.5 cm from base to apex. Two white standard triangles measured respectively 8.5 cm and 6.5 cm in height. One at a time, these standard triangles were presented approximately at *S*'s eye level and directly to his left, while the movable comparison object remained directly in front of him at the same height as the standard. Lines projected upon *S* from the two stimuli would thus have made an angle of 90°. Between *S* and the comparison object was interposed a row of chairs, which reached a height well below *S*'s eye level. No objects were interposed between *S* and the standard triangle.

---

[1] The present investigation was designed and the data obtained when the author was at Harvard University.

Daylight was excluded, and the room was artificially lit for the experiment.

Subjects

The 40 men who served in the experiment were all students of Harvard College or of one of the Harvard graduate schools. All professed normal vision or vision corrected to normal by means of eyeglasses.[2] Ages of the *S*s ranged from 18 to 24, with a median of 20.5 years. The selection of a poorly adjusted group, numbering 20 cases, was made by an experienced clinical psychologist,[3] on the basis of referrals to the Harvard Department of Hygiene and from his own psychological examination of the cases. Various difficulties of adjustment characterized these persons and, in general, it may be said that most of them suffered from some form of psychoneurosis. The control group of 20 cases was recruited by *E* through solicitation of the services, with the offer of remuneration, of Harvard men randomly chosen. These persons were also examined by the clinical psychologist and found apparently free from symptoms of poor adjustment.

Procedure

A test battery, including Guilford's Inventory of Factors STDCR (6), the Reasoning subtest from the Chicago Tests of Primary Mental Abilities (14), and Thurstone's Concealed Figures Test (13), was individually administered to each *S* in the first of two sessions. The size-distance judgments were obtained in the second session, held on a subsequent day. The first judgment was obtained with the comparison triangle 30 ft. in front of *S* and the 8.5 cm standard triangle 15 ft. to his left. Instructions were of the "objective" type; i.e., *S* was asked to make the matches with reference to the physical size of the stimuli and to give the judgment "equal" only when he thought the two would measure the same amount. One ascending and one descending series were administered under these conditions. The standard was then placed at a distance of 2 ft. from the bridge of *S*'s nose when *S*'s head was turned directly to the left, and two more judgments were obtained. *S* and the standard triangle were then both moved to a position where *S* was 15 ft. in front of the comparison object and the standard was 12 in. to his left. After *S* had made two more judgments, the standard was moved to a position 15 ft. to his left, thus making the stimuli equidistant from him; two further judgments were then obtained.

S returned to his seat at the 30-ft. distance, and the 6.5 cm standard was placed at the 15-ft. position. *S* was now given new instructions, this time of the "analytic" type. He was asked to ignore distance and what "knowledge" he might have of the sizes. His task was to select from the distant comparison the size which would be just covered by the standard triangle if the latter was suspended in his forward line of vision and at its present distance. This definition of "retinal" equality was further illustrated by reference to the areas of space occupied by the two triangles if seen together through an hypothetical window frame at a fixed distance in front of *S*. *S* then made four more pairs of judgments at the same distances and in the same order as for the previous set of objective judgments.

RESULTS

Each major group (poorly adjusted and normal) was split at its median on the R scale of Guilford's Inventory of Factors STDCR. Since this scale is claimed to be one of the best available measures of introversion-extraversion (4, 5), it was supposed that the above procedure would isolate four subgroups which may be respectively styled "neurotic introverts," "neurotic extraverts," "normal introverts," and "normal extraverts." As it happened, the obtained range of R scores extends more towards the extreme introversion end of the scale than towards the extreme extraversion end. Strictly, therefore, the extravert groups should be given the designation "relative extravert." Table 1 shows for each subgroup the mean and *SD* on the R scale, and also indicates the means of the size matches for the various distances and the two conditions of judgment.

Preliminary analysis showed heterogeneity of variance among the measures of size-distance judgment, and they were therefore given a logarithmic transformation. Separate analyses of variance for analytic and objective

TABLE 1

EXPERIMENTAL CONDITIONS, MEAN SIZE MATCHES IN CM AND MEAN R SCORES

Instructions, standard size	Distances (ft.)		Groups			
			Neurotic (N = 20)		Normal (N = 19)	
	Comparison	Standard	Introv.	Extrav.	Introv.	Extrav.
Objective, 8.5 cm	30	15	8.45	8.25	8.40	9.08
	30	2	7.48	7.72	7.78	8.25
	15	1	7.52	8.10	8.37	8.47
	15	15	8.49	8.70	8.75	8.83
Analytic, 6.5 cm	30	15	9.93	8.17	8.08	8.47
	30	2	22.40	17.45	14.27	13.97
	15	1	23.90	18.47	18.24	14.21
	15	15	6.60	6.49	6.50	6.83
R score:						
Mean			19.70	32.30	26.90	43.20
SD			5.22	7.10	5.13	7.00

---

[2] After completion of the experiment it was found that one *S* had incorrectly made this claim and in fact suffered from uncorrected strabismus. His size-distance judgments were therefore excluded from the analysis.

[3] The author is indebted to C. C. McArthur of Harvard University for examining the cases and making the diagnoses.

conditions respectively were performed on these data and the results are shown in Table 2. It will be seen that for analytic conditions "neuroticism" is a significant source of variance, neurotics more than normals tending to match the visual angle. The F value for introversion-extraversion is not large enough to warrant rejecting the null hypothesis. It must be remembered, however, that the groups were not selected for introversion-extraversion as they were for neuroticism. Examination of Table 1 shows that the two traits are not entirely independent, neurotics tending to have lower R scores than normals.

Comparison of the different pairs of subgroups yields clarification. Analysis of variance of the analytic judgments by the two neurotic subgroups yielded an F value of 4.11 with 1 and 18 *df*, which value just misses significance at $P = .05$. The tendency indicated is for introverts to make their matches more in the direction of visual angle than in the direction of object size, and vice versa for the extraverts. Similar analysis of the two normal groups revealed no such tendency ($F = 0.08$). Only one pair of groups revealed an unequivocally significant difference—neurotic introverts and normal extraverts. The F value for between-group variation is 10.09 with 1 and 17 *df*, $P < .01$, and for the Distances by Groups interaction is 6.05 with 2 and 34 *df*, $P < .001$. Neurotic introverts thus make matches more in the direction of visual angle and are more responsive to changes in distance conditions. The matches of the normal extraverts tend toward or beyond the object size and tend more to remain stable despite changes in distance conditions.

TABLE 2

ANALYSES OF VARIANCE OF SIZE MATCHES

	df	Analytic		Objective	
		Mean Square	*F*	Mean Square	*F*
Neuroticism (N)	1	.4563	6.32**	.0152	2.49
Introversion (I)	1	.1658	2.30	.0085	1.39
N × I	1	.0529	0.73	.0004	0.07
Subjects[a]	35	.0722		.0061	
Distances (D)	2	.9601	72.73****	.0161	7.67***
N × D	2	.0573	4.34*	.0007	0.33
I × D	2	.0223	1.69	.0001	0.05
N × I × D	2	.0061	0.46	.0039	1.86
Residual	70	.0132		.0021	

* $P < .03$.

** $P < .02$.

*** $P < .01$.

**** $P < .001$.

[a] Since one S was lost from the normal extravert group (see text, footnote 3) his scores were replaced by the appropriate cell means. One *df* was then subtracted from the "Subjects" classification and 3 *df* were subtracted from the total.

It is clear, therefore, that while neuroticism appears to be the major source of between-group variation, it must be admitted as a distinct possibility that introversion is a contributing factor in analytic size-distance judgments.

Analysis of variance of the objective judgments, also summarized in Table 2, shows no significant differences between groups and no significant interactions. The tendency present, under these conditions, is opposite to that for analytic judgments, normal persons making responses slightly more in the direction of visual angle than the neurotics. The latter, at the disparate distances, uniformly showed "overconstancy."

The results described above are obtained with adequate control of factors previously shown to be possible functions of size-distance judgment, namely, sex and age (12). A further possibility is that the significant difference in analytic size-distance judgments between neurotic and normal groups may be accounted for in terms of intelligence (12). Analysis of results from the two measures of intellectual ability yields no reason to believe that this is the case (Reasoning, $t = 1.02$, $P > .05$; Concealed Figures, $t = 0.85$, $P > .05$).

Data from the equidistant judgments provide a control on the possibility that the difference discovered is not due to distance but rather a constant error of size estimation such as found in other work on affective processes in perception (7). Inspection of Table 1 shows that there was an over-all tendency for all *S*s slightly to overestimate the size of the standard object at the 15-ft./15-ft. position. Analysis of variance of the data showed, however, no significant differences between the groups and no significant interactions.

DISCUSSION

The results of this experiment clarify, to some extent, the ambiguity resulting from studies by previous investigators. In the present experiment, poorly adjusted people behaved like Singer's "thinking introverts"

(11). As previously mentioned, Eysenck's factorial work (3, 4) suggests that Singer's so-called introverts actually are better described with reference to the neurotic dimension of personality. In the present experiment, the tendency of poorly adjusted persons to judge in terms of the visual angle was true only for analytic intructions, similar in intent to the "Look" instructions of Singer. Under objective instructions, these people showed an opposite tendency, i.e., toward constancy or "over-constancy," relative to the performance of normal persons. While this unexpected result does not reach statistical significance, it is interesting to note that it is in a direction consistent with the finding of Sanders and Pacht (10) who also used objective instructions. It is clear that no generalized statement can be made about the "tendency to size constancy" of psychoneurotic or poorly adjusted persons, without specifying the nature of the instruction-produced task attitude. Previous studies, together with the present results, suggest that the tendency is smaller than that of normal persons in the analytic situation and greater than that of normals in the objective situation.

The complexity of the problem is further realized when the role of introversion is considered. The present results make it difficult to resist the conclusion that neuroticism and introversion appear to be additive in their effect on analytic size judgment. Introverted neurotics differ markedly from extraverted normals in their analytic judgments, whereas normal introverts differ from neurotic extraverts only to an insignificant degree.

Singer (11) has adopted a position similar to that of Weber (15) in suggesting that the introvert, because of immersion in self and analysis of his own cognitive processes, is less dependent than the extravert on the stimulus characteristics of the external environment and therefore shows less tendency toward constancy. The present type of analysis shows that such a view is too simple. Introversion by itself was in the present experiment insufficient to produce the effect. Singer's reasoning is still plausible if it is made to refer to introverts who are also neurotic or, in other words, to Eysenck's "dysthymic" category. However, his explanation cannot account for the opposite tendency under objective conditions of judgment, found as a nonsignificant trend in this experiment, and found to a significant degree by Sanders and Pacht (10).

While there is danger in extrapolating from the present results to a different type of population, such a venture may be attempted in view of evidence and theory already existing in the literature. If we admit that the disease of schizophrenia represents an extreme both of introverted disposition and of maladjustment, it may be supposed that the victims of this malady would surpass even the introverted neurotics of the present study in their analytic tendency to match the visual angle. This suggestion may do violence to the widely held view of psychosis as a factor independent of neurosis, but it is consistent with the finding of Crookes (2) that neurotics have lower constancy indices than normals and that schizophrenics show indices significantly lower than either of the other groups. Also pertinent is Lovinger's (8) finding, cited earlier. In the meantime, therefore, the best theoretical rationale for the existing evidence seems to be an extension of Bruner's (1) suggestion that schizophrenia represents a withdrawal from object relations and therefore involves a breakdown in the constancies. Such a withdrawal from object relations may also occur to a lesser degree in neurotic introverts. The results inconsistent with such a theory may at present be attributed either to presentation of the size-judging situation as an objective rather than an analytic task (e.g., 10) or to a failure adequately to communicate the analytic nature of the task. Further study of the objective task is needed to yield understanding of why it has properties different from the analytic task in relation to adjustment and disposition.

SUMMARY

A size-distance judging task was given to 40 men who were classified in four groups, respectively styled "neurotic introverts," "neurotic extraverts," "normal introverts," and "normal extraverts." Analysis of data from four distances under two conditions of judgment, i.e., objective and analytic, indicated that neuroticism was the major source of between-group variation. Under analytic conditions, neurotic persons tended to match the stimulus in terms of visual angle, and normals in terms of object size. The data also

indicated that introversion to a less marked degree was also related to matches in the direction of visual angle. Objective judgments showed the opposite trend to a nonsignificant degree. The results were discussed in relation to previous research in which the several variables of the present design were separately treated with some inconsistent outcomes.

REFERENCES

1. BRUNER, J. S. Personality dynamics and the process of perceiving. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: An approach to personality.* New York: Ronald, 1951. Pp. 121–147.
2. CROOKES, T. G. Size constancy and literalness in the Rorschach Test. *Brit. J. Med. Psychol.*, 1957, **30**, 99–106.
3. EYSENCK, H. J. *The scientific study of personality.* New York: Macmillan, 1952.
4. EYSENCK, H. J. *The structure of human personality.* New York: Wiley, 1953.
5. FRANKS, C. M. Conditioning and personality: A study of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1956, **52**, 143–150.
6. GUILFORD, J. P. *Inventory of factors STDCR.* Beverly Hills, Calif.: Sheridan Supply, 1940.
7. JENKIN, N. Affective processes in perception. *Psychol. Bull.*, 1957, **54**, 100–127.
8. LOVINGER, E. Perceptual contact with reality in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **52**, 87–91.
9. RAUSH, H. L. Perceptual constancy in schizophrenia. I. Size constancy. *J. Pers.*, 1952, **21**, 176–187.
10. SANDERS, R., & PACHT, A. R. Perceptual size constancy of known clinical groups. *J. consult. Psychol.*, 1952, **16**, 440–444.
11. SINGER, J. L. Personal and environmental determinants of perception in a size-constancy experiment. *J. exp. Psychol.*, 1952, **43**, 420–427.
12. THOULESS, R. H. Individual differences in phenomenal regression. *Brit. J. Psychol.*, 1932, **22**, 216–241.
13. THURSTONE, L. L. *A factorial study of perception.* Chicago: Univer. Chicago Press, 1944.
14. THURSTONE, L. L., & THURSTONE, THELMA G. *Chicago Tests of Primary Mental Abilities.* Chicago: Science Research Associates, 1947.
15. WEBER, C. The relation of personality trends to degrees of visual constancy correction for size and form. *J. appl. Psychol.*, 1939, **23**, 703–708.

Received August 23, 1957.

VALUE DIFFERENCES AMONG SOUTH AFRICAN STUDENTS

KURT DANZIGER

University of Natal

ABERLE (1) has indicated that while "contemporary social theory . . . strongly emphasizes the crucial importance of a shared system of ultimate values as an element in any society" it tends to gloss over the diversity of subsystems that exist within any complex society. The South African situation is of interest in this connection as it provides an opportunity for studying value conflict under conditions of marked social tension.

The white minority in South Africa exercises a virtual monopoly of political, economic, and military power, and most of its elected representatives are committed to a policy of maintaining white supremacy at any cost. On the other hand, as the nonwhite majority is drawn increasingly into the web of a modern industrial civilization, its reaction to this state of affairs tends to assume a more and more determined character. The details of this conflict do not concern us here and may be gathered from a number of sources, e.g., Kuper (7, 8, 9).

Among the nonwhite section of the South African population the great majority are Bantu Africans, but there are also people of mixed descent and an Indian community which includes the most important nonwhite middle-class element. The white population is divided into English-speaking and Afrikaans-speaking sections.

In a homogeneous community the members share a common definition of their social situation, to use W. I. Thomas' term. They agree about their position in the world, their goals and how to achieve them; they have a similar evaluation of their society as a whole and of their position in it. In a society split by internal conflict this is not the case. The same social situation is defined very differently by the members of the opposing groups. What is regarded as a just and desirable social order by some is detested by others. The less agreement there is about the characteristics of the social situation in which the different groups find themselves the more intense the underlying conflict is likely to be. The first task of a study of value differences in a situation of social conflict is therefore the measurement of group differences in the definition of the social situation. In the South African context this means examining the evaluation that whites and nonwhites make of the dominant pattern of their society as it exists at present.

But the subjective manifestations of social conflict are not exhausted by differences in the definition of the social situation. They enter much more deeply into the personality and serve to determine the inner values of the individual in important ways. When people differ markedly in their conception of the society in which they live, they are also likely to differ significantly in the goals which they set themselves within that context. The individual who feels himself to be a member of an unjustly oppressed group is likely to have different personal aims and values than the individual who denies the existence of social oppression in his society. Differences in the definition of the external social situation thus involve concomitant differences about the role and duty of the individual in this situation. The second part of the present investigation is therefore concerned with assessing some of the main differences that arise between white and nonwhite South Africans in this respect.

METHOD

Subjects

The *S*s of the present investigation were 195 students at a number of higher educational institutions in South Africa. The total number was made up of 81 white university students, 34 white students at a technical college, 37 Indian and 43 African university students. The technical college students were only involved in the first part of the investigation. Of the white *S*s, 21 were Afrikaans-speaking. The socioeconomic level of the white group as a whole was considerably above that of the nonwhite group. The university students included in the sample were all attending introductory psychology or sociology courses in segregated classes and proceeding to arts, science or medical degrees. The technical college students were attending a course in pharmacy.

Questionnaires were administered on a voluntary basis in the classes attended by the *S*s. There was no consultation among individual *S*s.

Characterization of "White Civilization"

In the first part of the investigation, an attempt was made to assess the *S*s' over-all reaction to the social situation in South Africa by measuring their reaction to the key term "White Civilization." This term is frequently used in political and other discussions, both public and private, also in speeches and publications at all levels. Generally the context is one of a call for the preservation of "White Civilization." The term in fact refers to the ruling pattern of social relationships in the country. Thus, to ask people to characterize this entity is to ask them to define the dominant social pattern in their society. In this way it is possible to get an assessment of a key element in their definition of the social situation in which they live. In a different national context, the same purpose might have been served by getting characterizations of the meaning of such terms as "democracy," "free enterprise," "the American Way of Life," etc.

*S*s were presented with a list of 14 features which, according to the instructions, "different people have claimed to be highly characteristic of white civilization in South Africa." They were asked to indicate which of these features they considered to be really characteristic of white civilization and which not. They were also asked to rank the former in order of importance. The items available for choice were:

1. High standards of morality in the sphere of family life
2. Respect for law and order
3. Unjust oppression of nonwhite people
4. Great scientific achievements
5. A superior system for the production of material goods
6. Exploitation of cheap black labor
7. Arbitrary and undemocratic government
8. An unfair distribution of material goods
9. An emphasis on the rights and liberties of the individual
10. The ideal of equality of opportunity for all
11. Brutality towards the underdog
12. Hypocrisy in dealing with nonwhites
13. A legal system based on justice and fairplay
14. The ideal of peace among all men

Personal Values

On the hypothesis that differing evaluations of the social situation would be associated with preferences in respect to individual goals, *S*s were asked to answer the following open ended questions:

1. What two things would you like to have that you don't now have?
2. What two things would you most like your child to have?
3. What two things could you conceivably accomplish during your lifetime that you would be most proud of?
4. For what end would you be willing to make the greatest sacrifice of personal confort, time, and money?
5. What are the two worst things that could happen to you during your lifetime?
6. As a parent, what two specific lessons will you try hardest to teach your children?

The goals mentioned in reply to the first four questions were categorized as follows:

Political goals: e.g., "a new system of government," "the right to vote," "freedom of speech and movement," "the right to choose how I am governed," "political freedom," "liberation of my people," etc.

Educational goals: in the sense of a desire for improved educational opportunities, not of personal success

Specific helpfulness: e.g., "help my people," "give medical service to poor Africans," "establish a clinic in an African area," etc.

Abstract helpfulness: e.g., "reducing human unhappiness," "help those in need," "creating peace and universal goodwill," etc.

Familial goals: e.g., "a happy family life", "raising a family" etc.

Material possessions: e.g., "money," "a house," "a motor car," etc.

Personal characteristics, exploits and achievements: e.g., "ambition," "shining at sports," "getting my degree," "health," etc.

The categories used in analyzing responses to the last two questions were somewhat different and are discussed below.

Abbreviated F Scale

An abbreviated form of the F scale was administered to the white university students only. This consisted of Items Nos. 1, 4, 8, 9, 18, 19, 26, 27, 29, 31, 34, 37, 42 selected from the list of items (Forms 45 and 40) presented by Adorno et al. (2, pp. 255–257). Each item was scored on a five-point scale, a score of 0 representing strong disagreement and a score of 4 representing strong agreement.

Results and Discussion

Characterization of "White Civilization"

The 14 choice items presented as possible characteristics of "White Civilization" actually include eight items (Nos. 1, 2, 4, 5, 9, 10, 13, 14) that refer to favorable characteristics and six items (Nos. 3, 6, 7, 8, 11, 12) that refer to unfavorable characteristics. The proportion of favorable to unfavorable items in the choice of each *S* may be taken as an over-all measure of his reaction to the term "White Civilization." In assessing this proportion, only the first five choices for each *S* were taken into account. By means of this procedure, *S*s may be readily divided into three groups: (*a*) Those who choose not more than one negative item in their first five choices—these *S*s are clearly favorably disposed to the social situation designated by the stimulus term; (*b*) those *S*s who choose not more than one positive item in their first five choices and whose reaction is therefore unfavorable on the whole; (*c*) those *S*s who choose more than one positive *and* more than one negative item and who may be regarded as being intermediate in their reaction to the stimulus term. The frequency of the three response types for each of the experimental groups is shown in Table 1.

TABLE 1
Frequency of Different Over-all Responses to "White Civilization"

Response Category	White[a]		Nonwhite[a]	
	University Students	Technical College	Indians	Africans
Favorable	22	23	5	3
Intermediate	27	6	9	5
Unfavorable	23	5	23	33
Total	72	34	37	41

[a] Nine white university students and two Africans returned the questionnaire unanswered.

The table shows (*a*) that whites tend to evaluate the stimulus more favorably than nonwhites; (*b*) that the proportion of favorable evaluations is much less among white university students than among technical college students; and (*c*) that the proportion of unfavorable evaluations may be slightly greater among African than among Indian *S*s. These differences can be related to marked differences in the social situation of the different groups. The white *S*s as a whole tend to evaluate favorably a social order of which they are the beneficiaries, while the nonwhite *S*s evaluate negatively an order that reduces them to the level of second-class citizens. On the one hand, the situation is defined primarily in terms of certain positive characteristics; on the other hand, the situation is seen as threatening and repressive.

The difference between white university and technical college students and the unusually high proportion of "intermediate" response types among the former may be a reflection of a conflict that exists for the university students between the stereotyped attitude of their homes and a more critical attitude to which they are exposed at the university. In this and other respects the technical college students are much more representative of the white population as a whole; an analysis of white voting behavior certainly indicates an overwhelmingly positive evaluation of "White Civilization."

The difference between African and Indian students may be a function of the objective fact that under the existing social order the former have even fewer civil rights than the latter.

Nature of Favorable Characteristics Chosen

Further light is thrown on the over-all difference between white and nonwhite *S*s by an analysis of the *nature* of the favorable items that are chosen by each group as being characteristic of "White Civilization." The most popular items for both white groups are Nos. 1, 2, and 13—"high standards of morality in the sphere of family life," "respect for law and order," and "a legal system based on justice and fairplay." It is noteworthy that all three items, and particularly the first and the last, would appear to establish some claim as to the *moral* advantage of "White Civilization." On the other hand, insofar as nonwhite *S*s choose any favorable items at all, these tend to be items concerned with the *material* advantage of "White Civilization," namely, Items 4 and 5, "great scientific achievements" and "a superior system for the production of material goods." The difference between the groups in this respect is presented in Table 2.

Thus, while the order of which they are the privileged representatives appears to the white *S*s as moral and just, the nonwhite *S*s are prepared to concede at most material advantages

TABLE 2

NUMBER OF CHOICES OF FAVORABLE "MORAL" AND "MATERIAL" CHARACTERISTICS OF "WHITE CIVILIZATION"

	White	Nonwhite
Choice of "moral" characteristics, Items 1 and 13.	67	9
Choice of "material" characteristics, Items 4 and 5.	15	21

Note.—$\chi^2 = 27.5$; $p < .01$.

TABLE 3

RESPONSES TO THE QUESTION: "WHAT TWO THINGS WOULD YOU LIKE TO HAVE THAT YOU DON'T NOW HAVE?"

Response Category	White	Nonwhite
Political goals	Nil	61
Educational opportunities	4	15
Familial goals	33	5
Personal qualities and achievements	69	16
Total	106	97

to the same order, rejecting entirely its claims to moral excellence. Thus the latter group defines the dominant social pattern as basically immoral and unjust. The items it most frequently chooses as characteristic of "White Civilization" are Nos. 3, 6, and 7, "unjust oppression of nonwhite people," "exploitation of cheap black labor " and "arbitrary and undemocratic government."

Certain differences of attitude within the white group may be mentioned briefly. While the English-speaking *S*s often accept at least one unfavorable item as being characteristic and reject a few favorable items as not being characteristic of "White Civilization," the Afrikaans-speaking *S*s show an extremely rigid tendency always to accept only favorable characteristics and to reject only unfavorable characteristics. For these *S*s therefore, the term "White Civilization" acts as a true stereotype towards which no conscious ambivalence can be tolerated.

A difference that arises between English-speaking university and technical college students relates to the area in which negative characteristics are admitted. Both English-speaking groups tend to accept Item 6—"exploitation of cheap black labor," and to reject Item 10—"ideal of equality of opportunity for all." But only the university students, to any extent, accept Item 7—"arbitrary and undemocratic government"—as characteristic, and reject Item 9—"an emphasis on the rights and liberties of the individual" as not characteristic of "White Civilization." This finding suggests that while most English-speaking *S*s are prepared to concede the existence of a certain amount of unfair discrimination in the economic sphere, only the university students are prepared to extend criticism to the political sphere. Insofar as they do this they are presumably applying democratic standards that are foreign to the ordinary white citizen.

Personal Values

The second part of the investigation was concerned with differences of personal values between white and nonwhite groups. For this part of the study, the technical college students were not available, and the results are therefore based on university students only.

Replies to the open-ended questions were scored in accordance with the categories described earlier. The frequency of all the response categories among white and nonwhite *S*s was tabulated separately for each of the six questions. In each case a value of χ^2 was obtained that was significant at the .01 level. Large and consistent differences were found to exist between the two groups. Table 3 shows the main differences that occurred in the responses to the first question.

The difference between the two groups is striking. Whereas the white group concentrates its choices in the sphere of purely personal satisfactions, the nonwhite group tends to put wider social interests first. The same pattern of group differences occurs with the responses to each of the six questions, and in each case, the difference is highly significant statistically. The distinction between the response pattern of white and nonwhite groups appears to be similar to one made by Allport and Gillespie (4) who use the term *privatism* to refer to the American and often the European student's concern with "private values" and his absorption in the "goals of private worldly success." By contrast, Egyptians, Mexicans, Japanese, and South African Bantus were found to be far more orientated towards political and social issues. Similarly, Hughes and Thompson (6) in the United States found that Northern white children put a lower value-emphasis on justice and group unity than Southern Negro children. Lee (10) and others have stressed the basic nature of individualistic values in the culture patterns of societies generally designated as "Western."

One reason for this difference probably lies in the fact that certain societies have a strong tradition of economic individualism, whereas others do not. But the sociopolitical orientation of certain groups is not simply negatively conditioned by an absence of individualistic values; it is a result of certain positive pressures towards redressing real and perceived limitations on the group by means of group action. Members of national and class groups whose aspirations and "path of free movement" are limited simply because of this group affiliation tend to internalize the social aspirations of the group so as to turn them into individual aspirations for each member. In the case of less privileged groups, social and individual goals tend to coincide, whereas in dominant social groups, conflict between public duty and individual interest tends to occur. While most of the nonwhite *S*s who evaluated the existing dominant social order negatively also mentioned a change in this order as one of their personal desires, not one of the white *S*s did so, although 23 of them had apparently recognized the injustice of this order. For these *S*s there exists a characteristic discrepancy between the definition of the social situation and the personal aspirations of the individual.

Further light is thrown on this pattern by comparing the frequency of the two response categories—"abstract helpfulness" and "specific helpfulness"—for Questions 3 and 4. Table 4 combines responses in these two categories for both questions. It is in accordance with the whites' basically individualistic orientation that when they do declare themselves to be willing to make sacrifices for an end other than their private satisfaction, this end tends to assume a conveniently abstract and nebulous character.

TABLE 4

FREQUENCY OF "HELPFULNESS" RESPONSES

(Questions 3 and 4)

Response Category	White	Nonwhite
Abstract helpfulness	29	15
Specific helpfulness	6	30

Note.—$\chi^2 = 20.6$; $p < .01$.

On the whole, the person with a specific social aim, like building a clinic, for example, is much more likely to translate his aim into social action than the person who aims merely to "reduce human unhappiness." The point is that social action must always be concrete and the person with an abstract aim must first give his aim a concrete content in terms of the situation that actually faces him. As long as the aim remains abstract and formal, its function may not really be that of reorientating the individual towards social action, but rather that of assuaging the guilt that arises from the conflict between social ideals and private interest. For the socially oriented person, on the other hand, social aims naturally assume a concrete content, as they arise directly out of the demands of a specific external situation that have become identified with his individual interest. Thus the nonwhite *S*s repeatedly mentioned the goal of helping their own communities also in other than political ways. In the white *S*s, the split between private and public goals may be accompanied by guilt feelings where the individual is aware of the demands of the social situation and of his own failure to live up to them. Allport and Kramer's (3) finding that their less prejudiced *S*s showed guilt about what prejudice they did possess may be relevant here.

The pursuit of private satisfactions, however, is not only the result of the expectation of positive gains in that direction; to an extent which can never be accurately determined, it is also conditioned by fear of the consequences to the individual of participation in sociopolitical action. This factor varies with the level of social repression that exists in the society. In South Africa, for example, the individual faces serious penalties should he advocate a fundamental change in the existing social pattern. The socially oriented individual is therefore compelled to face the very real possibility of political persecution with all its attendant personal consequences for him. The privatistic individual escapes this, but as Table 5 shows, he may then develop less rational fears that are characteristic for his social isolation.

Before considering the important group differences shown in Table 5, it is worthwhile noting that physical misfortunes and ill health are mentioned with the same frequency by

TABLE 5
FREQUENCY OF VARIOUS RESPONSES TO THE QUESTION: "WHAT ARE THE TWO WORST THINGS THAT COULD HAPPEN TO YOU DURING YOUR LIFETIME?"

Response Category	White	Nonwhite
Political persecution	8	32
Economic insecurity	18	31
Physical misfortune, ill health	41	38
Loss of moral standing	9	18
Death of a relative	14	9
Mental instability	13	3
Loss or lack of love	36	5
Total	139	136

both groups, thus indicating that this response is not affected by the differential social forces acting on the two groups.

The main pattern of group differences follows the differences of orientation that have already been discussed. Members of the underprivileged nonwhite group, oriented as they are towards a change in the social situation, have to face political persecution and economic insecurity as real possibilities. But this eventuality is not only unpleasant in itself, it also threatens to undermine the dominant orientation of the individual towards social goals because of the possible consequences of such an orientation. Thus a number of the nonwhite *S*s also express a fear of losing their moral standing in their community by a failure to live up to their social ideals.

On the other hand, the white *S*s often mention certain unpleasant possibilities that are almost nonexistent in the nonwhite sample. Here we get such responses as "insanity," "losing one's faith in everybody and everything," "losing all one's friends," "being unable to love anybody," "being unhappily married," "not getting married," "loneliness," etc. These responses have been classified in the last two categories of Table 5. They all appear to indicate a basic preoccupation with social isolation. The privatistic individual pays the penalty of his withdrawal into a world of individual goals in the form of a basic anxiety regarding social isolation or "loss of love." As the social relationships of his private life are not based on common social aspirations but on the capacity of each partner to furnish the other with narcissistic satisfactions, they are subject to a basic instability. His real isolation would appear to make the privatistic individual hyperdependent on constant proofs of love from his surroundings, while at the same time his basically narcissistic orientation prevents him from forming relationships of a type that would ensure his real social integration.

Insofar as the privatistic individual's relationships to other people are not founded on common goals, there will be a certain tendency to stress the more external features of these relationships and to develop a social "persona" that is suitable for such superficial relationships. Some evidence in this direction is contained in the replies to the question: "As a parent, what two specific lessons would you try hardest to teach your children?" There is a certain tendency for the white students to more frequently state that they would teach such virtues as "good manners," "politeness," "sociability," "good behavior," "the ability to get on with others," as well as "sportsmanship," "cheerfulness," and the ability "not to show one's feelings." These responses occur 36 times in the white group and 14 times in the nonwhite group. This difference is at least suggestive of a somewhat greater concern with the *external appearance* of social integration in the white group. Presumably such concern would be related to the very real difficulties which the intensely privatistic individual faces in establishing genuine social ties.

Another type of response to the last question which is of special interest is that related to authoritariansim. The main responses here were "obedience," "discipline," and "respect." They occurred 15 times in the white group and 20 times in the nonwhite group. However, it would be a mistake to assign the same significance to these responses in the two groups. In the white group these responses occurred mostly among *S*s who had a favorable over-all response to "White Civilization" (difference significant at .05 level), but no such relationship exists in the nonwhite group. Moreover, when the F scale was administered to the white *S*s, it was found that those whose over-all response to "White Civilization" was favorable tended to get higher scores on the F scale (difference significant at .02 level). The product-moment correlation between *F* score and favorability of response to "White Civilization" (scored on a six-point scale) was .65. Thus in the white group, acceptance of exist-

TABLE 6
RELATIONSHIP BETWEEN F SCORE AND RESPONSES TO "WHITE CIVILIZATION"

Response to "White Civilization"	F Scale	
	Mean per item	*SD*
Favorable	2.29	.55
Intermediate	1.84	.58
Unfavorable	1.59	.63

ing social discrimination and ethnocentrism in some cases form a pattern with an emphasis on authoritarian values in the training of children and "fascism" as measured by the F scale that is reminiscent of the "authoritarian personality" (2).

But in the nonwhite group such authoritarian values are frequently associated with an intensely critical attitude to the existing social order. They are quite likely to arise from a rational concern with the necessity for group solidarity in the face of external social threats to its aspirations. When these *S*s were asked to explain the reasons for their social orientation, references to the necessity for group solidarity in the face of social oppression were common (5). The interpretation of the pattern of "authoritarianism" must always take into account the wider social context. In the present case, the position is further complicated by the survival of traditional patterns in the nonwhite group to an unknown extent. Some estimate of the influence of these patterns could be made by examining the differences that exist between Indian and African *S*s. This has not been done in the present study which is concerned with the over-all differences between whites and nonwhites. The differences within the nonwhite group were actually very much less marked than these over-all differences. To some extent the patterns of privatism and social orientation do arise out of the traditions of the groups concerned. But they are also conditioned by the objective social situation in which the groups find themselves at the present time and by their subjective definitions of that situation.

The relationship between the real situation and its subjective definition has a different significance for the privileged white and the unprivileged nonwhite group. Insofar as the latter recognizes the elements of oppression in the objective situation, and the great majority of the *S*s appeared to do so, it arrives at a definition of the situation which makes a fusion of individual and social interest possible.

On the other hand, members of the privileged group have two alternatives. They may identify self-interest with justice and approve of the existing social situation. As that situation is in fact based on naked power and the rule of the strong, their approval of it makes them predisposed to accept an openly Fascist ideology, which is what a number of our white *S*s with high F-scale scores have actually done. On the other hand, those white *S*s who possess standards of social life that render it impossible to identify self-interest with justice are affected by an internal conflict in which self-interest always wins, although at a certain cost. Of course it is possible for this conflict to be resolved by an identification with the oppressed, but no clear case of this occurred in our sample.

South African society is marked by an exceptionally severe conflict between the process of social integration fostered by a growing modern industrial system and the process of deepening caste divisions which is initiated by the power of a state that represents the dominant caste. This conflict leads to wide divergencies of outlook between the members of the privileged and nonprivileged castes. These divergencies of outlook make communication across caste barriers exceptionally difficult, as they appear to affect not merely "surface" opinions but also more basic aspects of personality organization. Thus disagreements about social values and individual goals exacerbate the basic conflict. For example, opportunities for limited personal advancement within the existing social framework are sometimes offered by official policy in accordance with its scheme of values only to be frequently rejected with indignation by the supposed beneficiaries whose own aspirations lie in quite another direction. The result is a further deepening of tensions, except in so far as counteracting influences are brought to bear on the system from outside.

SUMMARY

Groups of white and nonwhite South African students were required to choose features that they considered to be characteristic of "White Civilization" in South Africa. Great differences in their definition of the social situation in the country were indicated by the

fact that the nonwhite students attributed mainly negative features to the stimulus term while the white students responded to it much more favorably. The latter claimed certain moral advantages for "White Civilization," while the former regarded it mainly as a system of exploitation and oppression with at best certain material advantages.

This difference in attitude to the dominant pattern of social relationships in the country appeared to be linked to a difference of personal values. In response to a number of questions designed to elicit value choices, the nonwhite students put a great deal of emphasis on political values and also on the goal of helping their own communities in other ways. On the other hand, the white students were almost entirely limited to choices in the sphere of private satisfactions—they desired success in their career, in their family life, in the cultivation of certain personal qualities, and personal exploits. Where they did pay due to the value of general helpfulness this was done in a characteristically abstract manner that contrasted with the more concrete approach of the nonwhite subjects to social improvements.

These differences may be interpreted in terms of a basic value of privatism that determines the life organization of the white subjects and contrasts strongly with the social orientation of the nonwhite subjects. In the case of the latter, there is a certain identity of social and individual aspirations, whereas the privatistic individual is marked by a split between private and public goals with the former having much greater force for him.

The privatistic and the socially oriented individuals are subject to different forms of anxiety. Whereas the latter tends to fear political persecution and a possible loss of moral standing, the latter's great fear is of social isolation in its various forms. In some cases this seems to go with an emphasis on the external appearance of social relationships instead of on their genuine content. It may be suggested that the privatistic individual has a basically narcissistic orientation which stands in the way of his forming genuine social ties based on shared aspirations. At the same time he requires proofs of love from his environment. His basically exploitative attitude to people leads to the prevalence of fears of social isolation.

One way of dealing with this situation appears to consist in the pattern of "authoritarianism" where the individual identifies his personal interest with that of the most powerful group, and submission to the powerful becomes a basic value. There was evidence of the existence of this pattern in some of our white subjects.

REFERENCES

1. ABERLE, D. Shared values in complex societies. *Am. sociol. Rev.*, 1950, **15**, 495–502.
2. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.
3. ALLPORT, G. W., & KRAMER, B. M. Some roots of prejudice. *J. Psychol.*, 1946, **22**, 9–39.
4. ALLPORT, G. W., & GILLESPIE, J. M. *Youth's outlook on the future.* New York: Doubleday, 1955.
5. DANZIGER, K. Self-interpretations of group differences in values. *J. soc. Psychol.*, in press.
6. HUGHES, J. H., & THOMPSON, G. G. A comparison of the value systems of southern Negro and northern white youth. *J. educ. Psychol.*, 1954, **45**, 300–309.
7. KUPER, L. South African native: Caste, proletariat or race? *Soc. Forces*, 1949, **28**, 146–153.
8. KUPER, L. The control of social change: A South African experiment. *Soc. Forces*, 1954, **33**, 19–29.
9. KUPER, L. *Passive resistance in South Africa.* London: Jonathan Cape, 1956.
10. LEE, D. Are basic needs ultimate? *J. abnorm. soc. Psychol.*, 1948, **43**, 391–395.

Received August 23, 1957.

FAMILY PATTERNS CORRELATED WITH ADOLESCENT PERSONALITY STRUCTURE

ROBERT F. PECK

University of Texas

IN PRESENT day theories of personality development it is almost axiomatic that the emotional dynamics of the child's family play a dominant part in shaping his personality. Even today, however, the most detailed and persuasive evidence for this proposition comes largely from individual case studies and, chiefly, from clinical studies of maladjusted individuals, who most often seem to present a compelling reason for investigating their histories. There have been, of course, a sizable number of studies that relate certain overt traits of child behavior to similar traits in parents (1). Perhaps the most comprehensive effort to relate child behavior patterns to parental practices has been the Yale cross-cultural survey (2).

For the most part, these studies have adhered to overt behavorial characteristics that are readily observable in a common-sense way, both in children and in parents. Conceptual constructs about "inner" personality dimensions, on the other hand, have rarely been tested in any quantified way, or on more than a few individual cases. In short, most studies susceptible of rigorous hypothesis-testing have dealt with surface traits which psychodynamically oriented people sometimes regard with misgiving as an over-simplified correlating of symptoms, rather than a study of the persistent inner roots of behavior. Conversely, not many "depth" studies of personality structure have provided systematic measurements of child and family characteristics in samples large enough to allow mathematical tests of significance; and where such work has been done, it has generally been on "problem" samples, such as delinquents (3).

METHOD

Sample

In the course of an eight-year longitudinal study of adolescent character development, data became available that offered an opportunity to measure certain key dimensions of personality structure, and to relate them to fairly universal dimensions of family experience in a quantitative way. The sample population represented a relatively even cross section of all the children born in the year 1933 in a sociologically typical mid-western town, known here as "Prairie City." The intensity of the study limited the sample to 34 children (half girls and half boys) and their 34 families, but at least some tests could be made of some of the salient propositions now current about the influence of family constellations on dimensions of personality structure. Indirectly, the data permit a test of the validity of such constructs as "ego strength," "superego strength," and the like.

The population consisted of 34 children, selected from the 120 children living in Prairie City in 1943 who had been born there ten years before. The original selection was based on a composite assessment of personal adjustment, comprised of teacher-ratings, test scores and clinician ratings. A pilot study indicated that the 34 children were quite evenly distributed along a scale from the best adjusted to the most poorly adjusted in the total group of 120. (As a bit of parallel evidence, they were evenly distributed, also, from top to bottom of the total group on a composite rating of moral character reputation, based on peer and teacher ratings.)

The Research Data

Each child was intensively observed, tested, and otherwise studied from age 10 to age 18. There were annual interviews by resident field workers with the child, his or her parents, and with friends and other informants. Sociometric tests of varied kinds were obtained annually, as were projectively analyzable essays. The Rorschach was administered at 13 and 14; the TAT at 13, 15, and 17; and a Sentence Completion test at 16. Numerous intelligence tests were repeated each year or two. A variety of self-rating questionnaires on values, attitudes, personal adjustment, family adjustment, etc. were obtained annually. Teacher ratings of several kinds, anthropometric measures and certain other data were also gathered.

Procedure

In 1946 a research staff of 12, under the direction of William E. Henry and the late Caroline Tryon, was engaged in a study of the adolescent role behavior of the children in the sample. Interested in the developmental backgrounds of the subjects, the staff, among other things, rated the children's families on eight aspects of their emotional relationships and child-rearing practices. These data were subsequently filed away and remained unexamined until the work of the 1950 conference had been completed.

In 1949–50, a new staff of ten, under the direction of Robert J. Havighurst and the writer, conducted a study of the character development of the same 34 adolescents. An intensive case study was made of each child. Among other things, the staff made ratings (as of the subjects' sixteenth year) on some 30-odd personality variables and, also, on two aspects of parental discipline (4). The estimated—Spearman-Brown—reliability of the judges' pooled-average ratings for any subject on any trait was .96.

Thus, two independent sets of data were available: one on the characteristics of the children's families, and the other on the personalities of the children. (The two additional family ratings by the 1950 staff were retained for analysis, although these were not strictly independent of the personality ratings, being made by the same judges.) A quantitative comparison of the measures of family dynamics with the personality dynamics of the children supplies the basis for the present report.

RESULTS AND DISCUSSION

Dimensions of Family Interaction

The ten family variables were intercorrelated and factor analyzed. Four trait-clusters were defined, with the following composition:[1]

- F1 Consistency in Family Life
  - Regularity of home life
  - Consistency of parental control
  - Common family participation in activities
- F2 Democracy–Autocracy
  - Child's degree of sharing in family decisions
- F3 Mutual Trust and Approval
  - Good father–mother relations
  - Parental trust and faith in the child
  - Child's willingness to share confidences with parents
  - Parental approval of the child
  - Parental approval of child's peer activities
- F4 Severity–Leniency
  - Severity of parental control

Dimensions of Adolescent Personality Structure

Twenty-nine of the personality variables rated by the 1950 conference were intercorrelated and factor analyzed (4). Six distinct trait clusters emerged as follows (the traits in each cluster are listed in descending order of their factor-loadings):

- P1 Ego Strength[2]
  - Emotional maturity
  - Locus of concern (egocentric to altruistic)
  - Internal consistency (roughly, personality integration)
  - Rationality of behavior
  - Accuracy of self-perception
  - Accuracy in assigning responsibility (to self or others)
  - Accuracy of social observation
  - Functioning intelligence (largely, test–IQ)
  - Insight into others' motives
  - Autonomy (inner-directed)
  - (Positive) Relations with the same-sex peers
  - Positive feeling toward mother
- P2 Superego Strength
  - Superego strength (presence of an effectively behavior-guiding conscience)
- P3 Willing Social Conformity
  - Overt conformity to the (conference defined) moral code
  - Emotional stability (largely, overt conformity to expected, "controlled" behavior)
  - Range of moral horizon (the range of people and groups toward whom *S* is overtly moral)
  - Overt acceptance of mother's expectations
  - Absence of overt hostility
  - (Positive) Relations with opposite-sex peers
- P4 Spontaneity
  - Identity of behavior with impulse
  - Positive feeling toward father
  - Empathy (with others' feelings)
  - Positive feeling toward same-sex peers
  - (Lack of) Guilt about overt behavior
- P5 Friendliness
  - Absence of covert hostility
- P6 Hostility–Guilt Complex
  - Guilt about inner impulses
  - Inner feelings of hostility toward mother
  - (Lack of) Overt acceptance of father's expectations
  - (Lack of) Inner liking for opposite-sex peers

Relationships of Personality Dimensions to Family Characteristics

In order to compare the two sets of data, each child's standard score on each of the six personality dimensions was computed. Similarly, a standard score was derived for each family on each of the four family dimensions. It was thus possible to correlate the children's scores on the personality dimensions with the characteristics of their families.

From Table 1, the significant correlations suggest the following relationships, presumably causal in nature, between family experience and personality:

Ego Strength seems largely produced by Trust (.73) and Consistency (.56).

Social Conformity also seems importantly influenced by Trust (.60) and Consistency (.53).

Superego Strength seems influenced by Consistency (.50) and perhaps Trust (.33).

---

[1] See Table 1 for intercorrelations of these four trait clusters. These were factorially distinct clusters, not "pure," uncorrelated factors. See appendix of *Psychology of Moral Character* (4) for the rationale for this use of what Stephenson calls "mixed" factor types or clusters.

[2] See Table 1 for intercorrelations of these six trait clusters. The factoring was done in order to reduce the original, overlarge set of variables to a small number of more stable variables, with the result that six distinctly separate traits or trait clusters were found, defined by a three-factor system.

TABLE 1

CORRELATIONS OF FAMILY AND PERSONALITY VECTORS

Variable	F1	F2	F3	F4	P1	P2	P3	P4	P5	P6
F1 Consistency in Family Life		−.01	.66**	.37*	.56**	.50**	.53**	.05	.19	−.10
F2 Democracy–Autocracy			.53**	−.65**	.43*	−.07	.16	.36*	.33	−.40*
F3 Mutual Trust & Approval				−.22	.74**	.33	.60**	.27	.44**	−.40*
F4 Parental Severity					−.16	.26	−.08	−.38*	−.38*	.40*
P1 Ego Strength						.48**	.73**	.56**	.43*	−.33
P2 Superego Strength							.66**	−.24	.02	−.23
P3 Social Conformity								.11	.48**	−.37*
P4 Spontaneity									.39*	−.28
P5 Friendliness										−.75**
P6 Hostility–Guilt Complex										

* Significant at .05 level.
** Significant at .01 level.

Spontaneity seems influenced by Democracy (.36) and perhaps Trust (.27), plus a lack of Severity (−.38).

Friendliness seems most influenced by Trust (.44) and Democracy (.33).

The Hostility-Guilt Complex seems influenced by Severity (.40) combined with a lack of Trust (−.40) and of Democracy (−.40).

The over-all pattern shown in Table 1, suggests that mutual trust among family members (F3) is rather closely related (.66) to consistency in family life (F1). It might be inferred from the correlation pattern that these two family characteristics, working together, are responsible for the development of healthy *ego strength* (P1) and willing *social conformity* (P3). That is, a child's emotional maturity, personal integration, autonomy, rationality of behavior, and his willingness to adapt to society's expectations, appear to be directly related to the degree of consistency, mutual trust, and mutual approval he experiences within his family.

Superego strength (P2) is not unrelated (.48) to ego strength (P1), but insofar as it is a separate entity, it appears to be the product of a somewhat different pattern of these family forces. It seems mainly to come from a regular, consistent family life. Superego strength is not at all the same thing as the hostility—guilt complex (P6); indeed, they correlate slightly negatively, −.23. Moreover, superego strength is unrelated to the generalized attitude of friendliness (−.02). One might therefore think that superego strength would be associated with severe parental autocracy (F2, F4); but it is not. Judging from the case studies, the reason appears to be that while some children with strong, *rigid* superegos do tend to come from severe, autocratic families, there are other children who have equally strong internal principles but who test and apply them rationally. These latter children came from families that have been quite democratic and unsevere. Other data from the study suggest the existence of these two different *kinds* of strong superego, each of which has a different relationship with the democratic-autocratic family dimension. Spontaneity (P4) and ego strength (P1) are moderately related. Both characteristics appear to be favored by a lenient, democratic family atmosphere which actively promotes a self-accepting, good-natured spontaneity in the child.

Friendliness is significantly related to the degree of mutual trust and affection within the family, whether or not there is stable consistency in it. There are some friendly children who have weak egos and weak superegos. These come (in the present study, at least) from inconsistent, irregular families who follow a policy that might better be called laissez-faire than "democratic," though the parents show an uncritically acceptant attitude toward the children in a desultory, rather inattentive way.

The hostility-guilt complex (P6) might readily be called a hostile but dependent, unresolved Oedipal complex, upon examination of the variables that comprise it (see above). Not surprisingly, it tends to occur in children whose families are severe, autocratic but unloving and inconsistent.

SUMMARY AND CONCLUSIONS

The purpose of the study was to test and measure the relationships between key elements in personality structure and certain characteristics of family emotional and regulatory patterns. Thirty-four adolescents were selected as a representative cross section of all children born in Prairie City in 1933. They were interviewed and diversely tested each year from age 10 to age 18.

Independent ratings were made on the families, and on the children, by two separate research staffs, following intensive individual case studies. The family ratings were intercorrelated and factor analyzed, yielding four dimensions of family interaction: Consistency, Democracy, Mutual Trust and Approval, and Parental Severity. The personality ratings, similarly analyzed, yielded six dimensions of personality: Ego Strength, Superego Strength, Willing Social Conformity, Spontaneity, Friendliness, and a Hostility—Guilt Complex.

Scores for each child were then computed on each personality dimension and these were correlated with the corresponding families' scores on each family dimension. The intercorrelations were analyzed, to discover general patterns of relationship between family experience and personality.

In this sample of adolescents, covering the full range of personal adjustment and moral responsibility, there proved to be a significant pattern of relationships between family experience and personality. Specifically:

1. Ego strength occurred in association with a family life which was characterized by stable consistency and warm, mutual trust and approval between the parents and between parents and child.

2. Superego strength was partially related to ego strength, but was chiefly related to the regularity and consistency of family life. It was *not* systematically related to severely autocratic rearing; or, at least, there appeared to be two different kinds of superego, produced in two different ways: (*a*) a strong, rigid, compartmentalized superego created by sternly autocratic rearing; and (*b*) a strong superego which was closely knit with ego functions, open to rational appraisal, and created by consistent, democratic, nonsevere rearing in a trustful, approving family. This interpretation rests, of course, on the qualitative case studies.

3. Generalized friendliness and spontaneity appear to be allied and to be associated with a lenient, democratic family atmosphere. Probably because the family background may or may not be consistent, these personality characteristics therefore may or may not be allied with ego strength and superego strength.

4. The hostility—guilt complex might reasonably be considered a hostile but dependent, unresolved Oedipal complex. It tends to occur in association with a severely autocratic, untrusting, and disapproving family.

In general, insofar as the findings of this study bear on the concepts of present day theory about personality development, they tend to corroborate that theory. These adolescents' personality characteristics proved to be significantly related to the emotional relationships and the disciplinary patterns which they experienced in living with their parents.

REFERENCES

1. SEARS, R. R., WHITING, J. W. M., NOWLIS, E., & SEARS, PAULINE S. Some child rearing antecedents of aggressions and dependence in young children. *Genet. Psychol. Monogr.*, 1953, **47**, 135–234.
2. WHITING, J. W. M., & CHILD, I. L. *Child training and personality.* Yale Univer. Press, 1953.
3. GLUECK, S., & GLUECK, ELEANOR. *Unraveling juvenile delinquency.* Harvard Univer. Press, 1950.
4. PECK, R. F. *The psychology of moral character.* Unpublished doctoral dissertation, Univer. of Chicago, 1951. (Microfilm)

Received August 26, 1956.

CASE REPORT

SOME DYNAMIC ASPECTS OF POSTHYPNOTIC COMPLIANCE[1]

MILTON J. ROSENBERG AND CHARLES W. GARDNER[2]
Yale University

In this paper we suggest, and offer case-record evidence in support of, a point of view that attempts to account for interindividual and intra-individual variability in compliance with posthypnotic suggestions. This point of view assumes that "depth" of hypnosis (if at all a meaningful concept) is neither the only nor the primary relatum of such compliance. Rather, the set of factors considered in this paper has to do with the content and phenomenal symbolic meaning to the subject of the posthypnotic suggestions that are directed to him.

THEORY AND HYPOTHESES[3]

It will serve our purpose to first regard the typical motives of the hypnotic subject and then to examine certain aspects of his relationship with the hypnotist. Our theoretical approach is based in part on a number of psychoanalytic contributions to the study of hypnotic phenomena (1, 2, 3, 4, 6, 7, 8).

The Subject

The public motivation of the experimental hypnotic subject (*S*) is for money, if any is offered, for "satisfaction of curiosity," and for association with a person of relatively greater status and power. These are derivatives of, and are overshadowed by, a need for a passive-receptive relationship with a person endowed with pseudo-magical attributes (6). The hypnotic *S* seeks a new person in response to the felt threat of loss of, or disappointment by, some significant figure in his life. That he is so motivated is often made clear through his expression of feelings of isolation, unrest, dissatisfaction, or even depression (7, 8). In the quest for a secure relationship, the *S* is often ready to forego some of his gratifications and his own untrusted strength, in exchange for the illusion of strength in the new object. He seeks to use the hypnotist as a partner in the acting out of a fantasy of oral fulfillment, ultimately a fantasy of the reunion of self with a parental image. At the same time he needs to use the hypnotist as a safe link to reality in this regressive dramatization. His exhibitionistic needs, too, are satisfied in the hypnotic relationship; particularly through the opportunity to relate exotic or remote material and to perform unusual hypnotic and posthypnotic tasks. Still other needs may be brought to the situation by the *S*, and these can generally be detected in his particular preconceptions about the nature of hypnosis.

The Trance

The "trance" is achieved ostensibly through the hypnotist's use of a formula emphasizing rhythmic repetition, monotony, and suggestions to relax (as well as other more specific suggestions) and by the *S*'s attempt to relinquish ostensible control or power and to submit himself to the hypnotist.

By means of the great diminution of sensory input and of channels of reality verification (through immobility, eye closure, etc.), to which the only exception is contact with the voice of the hypnotist, there occurs a fading or weakening of the boundaries of the self (4). The *S*'s need for magical participation with the hypnotist produces collaboration with this change, culminating in a partial merging of the images of the self and of a "sleep-giving" hypnotist.

In general, the contents of hypnotic imagery[4] and associations to it point to an underlying fantasy of oral or auditory incorporation that facilitates regression to a state of "fusion" between the images of the self and a powerful other (in this case the hypnotist).

Danger Points

Each hypnotic experience is unique, but most *S*s encounter certain anxiety-arousing "danger points." Initially, danger is felt in the terms of pretrance misgivings about surrendering control, and about the propriety of what the hypnotist may require.

---

[1] The authors wish to thank two of their colleagues, Roy Schafer and Irving Schneider, for their helpful evaluations of the original manuscript.

[2] Charles W. Gardner is in the Department of Psychiatry, Yale University.

[3] The theoretical approach proposed here is intended to apply primarily to the situation in which hypnotic experimentation is being done with volunteer *S*s. One would not expect an exact correspondence with the situation in hypnotherapy which is undertaken with *ostensibly* different motivations.

[4] We are indebted to G. O. Morris for unpublished observations on the function of imagery in hypnosis.

Somewhat later in the process, anxiety about compliance appears in images or thoughts of sexual attack, mutilation, or castration. The more profound dangers accompanying further progress in hypnosis are the fears associated with orality; fears of extinction, of being swallowed up or bitten, of swallowing up the hypnotist or destroying him, and fears of aloneness and loss of the self or of others.

The *S*'s past history and current psychological state determine which, if any, dangers he experiences and how he will respond to these dangers. The anxiety aroused in these situations can lead to a reversal or "lightening" of the hypnotic process; or the *S* may maintain himself in hypnosis through the use of characteristic mechanisms aimed at inner avoidance of anxiety and at manipulation of the hypnotist's behavior or expectations. Improvement, with practice, in hypnosis and in following hypnotic suggestions, represents the gradual mastery of the anxiety arising in the danger situations and is achieved through rehearsal and stepwise resolution of the conflict.

Posthypnotic Suggestion

Posthypnotic suggestion introduces new and additional difficulties. It requires that the *S* continue the identification with the hypnotist while in a "waking condition." In this situation, two contradictory states must coexist; a state of identification with a gratifying, somniferent hypnotist; and an awareness of a real manipulating hypnotist. Another difficulty is that the posthypnotic suggestion itself can complicate the hypnotic relationship in various ways. It may, for example, put defenses against repressed impulses in jeopardy, as when the behavior called for would reverse a strongly reinforced reaction formation.

Typically, the presentation of a posthypnotic suggestion is accompanied by the suggestion that the *S* will have a temporary amnesia for all, or for a portion, of the hypnotic session. This suggestion in itself will frequently reawaken all the reluctance to surrender control that was first experienced in response to the original achievement of hypnotic trance.

Despite these various difficulties, many *S*s (even some who have not been fully successful in ordinary "trance") do manage to comply with posthypnotic suggestions and to manifest amnesia for them. But, as we have noted, the success of such compliance does frequently vary from *S* to *S* and, with regard to a single *S*, from one posthypnotic suggestion to another. It is in the light of this observation that we have formulated the question to which the remainder of this article is addressed: *what processes guide and govern the hypnotic subject's acceptance of, and adaptation to, specific posthypnotic suggestions?*

In the present experiment, the attempt was made to produce, through posthypnotic suggestion, changes in the *S*'s *attitudes* towards certain social issues or social objects. Such responses are not as easy to change as are more ego-peripheral behaviors. However, it is this very fact which makes them particularly useful in research that seeks to uncover the dynamics of posthypnotic compliance. For where change is difficult to achieve, the experimental hypnotic *S* will, through his motivated *striving* to achieve the change, reveal more about the dynamics of such adaptation than he would in situations where the suggested change is less challenging. Thus we would contend that the dynamic processes posited in the hypotheses stated immediately below operate not only in the case of posthypnotic suggestions of attitude change, but rather that they pertain to any posthypnotic suggestion that arouses some degree, however small, of resistance in the *S*.

Hypotheses on Posthypnotic Compliance

Hypothesis I. Compliance with a posthypnotic suggestion varies as a positive function of the extent to which the subject can interpret and manipulate its content into an acceptable form. This is accomplished through the employment of dynamic processes set into motion during hypnosis and used by the subject to maintain himself in hypnosis.

Underlying this hypothesis is the view that the posthypnotic suggestion, being an attempted manipulation or intrusion by the hypnotist, will be responded to in the context of the already existing atmosphere of the specific hypnotic relationship. A further implication of this hypothesis is that the fate of the posthypnotic suggestion will depend on its usefulness to the *S* in the dynamic state in which he finds himself as he strives to ward off the anxiety and other disruptions associated with his hypnotic experience.

The "dynamic processes set into motion during hypnosis" can be considered, for present purposes, to fall into two classes. These are: first, *the mechanisms* used by the *S* to deal with the imminent emergence of anxiety in the induction and trance; and second, the specific mixture of *affective reactions* generated toward the hypnotist and his aims throughout the relationship and crystallized during the trance.

In the first group, some of the commonly found mechanisms are: *denial*, initially of the hypnotist's presence or influence, but capable of extension beyond these to other perceptions; *projection*, involving for instance, the attribution to the hypnotist of unacceptable sadistic impulses; forms of *rationalization*, especially explanations of compliance which disregard the passive aims of the *S*; *intellectual distancing*, such as the device of adopting a detached observer stance toward one's own inner processes; *play*, involving a heightened emphasis on the make-believe and experimental nature of the hypnotic performance; *turning passivity into activity*, and thus mastering a traumatic situation by actively reproducing it upon the self. The ultimate device called into action if hypnosis is successful, is *identification* of the self with the image of a "good" (protective) hypnotist, or a "bad" (aggressive) hypnotist. The device of identification, while facilitating hypnosis, also satisfies one of the basic needs of the successful hypnotic *S*.

The mechanisms that have been listed do not necessarily operate singly and separately. They may combine and fragment; for instance, the illusory reversal of roles often reported by hypnotic *S*s depends, for its effectiveness, upon denial, projection, and identification, all of which may serve in mastering the threat of passive submission.

Hypothesis I suggests that the posthypnotic suggestion will succeed in the measure that its content allows the *S* to organize, interpret, and execute it in a manner consistent with and guided by the mechanism, or mechanisms, that he regularly uses to overcome the danger points of hypnosis. It also points to the significance of a second group of dynamic changes: the *S*'s more enduring reactions to the hypnotist and his assumed aims, and the affects mobilized by these reactions. For example, the hypnotist may be reacted to as a frustrating, gratifying, unfriendly, protecting, or aloof person. Thus an *S* smarting under what he feels to be arbitrary domination or ruthlessness by the hypnotist may respond to the posthypnotic suggestion in ways determined by his affect. He may view the posthypnotic suggestion as a further assault, another justification for resentment, or another opportunity for masochistic submission. According to Hypothesis I, the chances for successful performance of the posthypnotic suggestion will be enhanced if its content lends itself to organization and interpretation that are consistent with the affects mobilized in the *S* by his experience of the hypnotic relationship.

Any one *S* will employ one or more mechanisms to maintain himself in hypnosis and will experience some out of the spectrum of possible reactions to the hypnotist; it is the potential consonance of these mechanisms and reactions on the one hand with the subject matter of the posthypnotic suggestion on the other hand that will, according to our Hypothesis I, determine whether and to what degree the *S* complies with the suggestion.

Hypothesis II. Where a posthypnotic suggestion activates a continuing inner conflict, compliance with that suggestion varies as a function of the degree to which it facilitates safe discharge of the warded-off drive (or drives).

Virtually any posthypnotic suggestion that directs an experimental *S* to perform some response which under nonhypnotic conditions would be ego-alien, may thereby tend to activate some related area (or areas) of conflict. In our own experiment, the social attitude picked for change frequently represents the result of defensive restrictions against certain unacceptable impulses. Suggesting the reversal of such an attitude increases the pressure of the warded-off material (through the implied permission granted by the suggestion) and may thereby require added defensive operations. When the motivation for the maintenance of hypnosis is not strong these defensive operations may interfere with the trance or work against the execution of the posthypnotic suggestion. But more typically these defensive operations are affected and limited by the regressive concomitant of hypnosis.

Thus what is predicted in Hypothesis II is that in the situation in which the posthypnotic suggestion mobilizes an area of conflict, the typical hypnotic *S* will the more readily accept and execute the suggestion to the extent that it permits him safely to indulge and express previously warded-off but persistent impulses. This "safe" indulgence is made possible not only by the regressive state characteristic of hypnosis but is also facilitated by the *S*'s identification with the hypnotist whereby the hypnotist becomes a new institution within the *S*'s personality; as such, the hypnotist is experienced as a force which, *through* the posthypnotic suggestion, has the power to legitimize impulses which were previously experienced by the *S* as illegitimate or as endangering.

It is the impression of the authors that in situations in which the emotional problems and defensive tendencies of the hypnotic *S* can be studied, one or more of the processes postulated in our hypotheses will be found to govern the *S*'s success and style in complying with posthypnotic suggestions. In the study reported in the remainder of this paper, the available data permitted at least a partial test of the adequacy of our theoretical account.

METHOD

Eleven experimental *S*s were selected on the basis of their having demonstrated good potential hypnotizability. These *S*s were tested (with a questionnaire device) on their attitudes toward various social issues. Each *S* was then rehearsed in hypnosis to the point where he could carry out posthypnotic suggestions with amnesia for these suggestions. The number of sessions required to reach this criterion level varied from three to eight separate hours of hypnosis.

The *S*s were then individually run through an experimental session which consisted of a number of steps. First, each *S* took a test[5] designed to measure the *cognitive structure* of one of the attitudes that he had already expressed on the attitude questionnaire.

On this "cognitive structure test," developed earlier by one of the authors (5), the *S* works with 31 so-called "value cards." Upon each card is printed a different value term (i.e., a state, condition, behavior, or attribute that may be positively or negatively, strongly or weakly, valued by the *S*). Examples are: "all human beings having equal rights"; "people being well educated"; "making one's own decisions"; "giving expression to feelings of anger or hostility"; "complying with the wishes of persons in authority." The *S* first judges each of the 31 values in terms of its importance to him, using a scale with a range of 21 points. This scale runs from "gives me maximum dissatisfaction" (Category −10) through "gives me neither satisfaction nor dissatisfaction" (Category 0) to "gives me maximum satisfaction" (Category +10). The intervening categories represent the intervening degrees of dissatisfaction or satisfaction. In a second part of the test, he is required to judge these same value items in terms of whether, and to what extent, he thinks each value will be *attained* or *blocked* as a consequence of the attitude object. On this task he uses an 11-point scale running from "extreme blocking of the value through the influence of the attitude object" (Category −5) through "the value is neither attained nor blocked through the influence of the attitude object" (Category 0) to "extreme attainment of the value through the influence of the attitude object" (Category +5).

Thus after testing there are available for each value term, the *S*'s judgment of its importance to him as a value and his judgment of whether and to what extent that value will, as he sees it, be attained or blocked through the influence of the attitude object. These two judgments are algebraically multiplied for each value item respectively. By summing all the products, a single algebraic quantity is obtained representing the total import of the *S*'s pattern of beliefs about the influence of the attitude object upon attaining or blocking various states that he values to differing degrees. This "index of cognitive structure" was computed for each *S*.

Upon completion of the cognitive structure test, the *S* was hypnotized to a stable level. He was then given a suggestion of posthypnotic affect-reversal with regard to the attitude object upon which he had been tested for associated cognitive structure. For example, one of the *S*s who had expressed, on the attitude questionnaire, strong opposition to "living in Los Angeles" was given exactly the following hypnotic instructions:

> When you awake you will feel very much in favor of living in the city of Los Angeles. The mere idea of living in Los Angeles makes you feel happy and exhilarated. Although you will not remember this suggestion having been made, it will strongly influence your feelings after you have awakened. Only when the signal to remember is given will you be able to remember and only then will your feelings revert to normal.

The posthypnotic suggestion, then, was simply of a change in affect on some attitude issue. Nothing whatsoever was suggested about how the person would *think* or what he would *believe* on that issue. The specific experimental prediction was that *S*s who complied with the posthypnotic suggestion of affect-reversal would undergo a process of spontaneous cognitive reorganization modifying their beliefs about the value-attaining and value-blocking potentialities of the attitude object so as to rationalize the affect-reversal.

To test this prediction, the *S*, after being brought out of hypnotic trance, was retested with the attitude questionnaire device (to determine whether the posthypnotic suggestion of affect-reversal was being complied with) and then with the test of cognitive structure (to determine whether he had changed his beliefs in a way which served to rationalize the affective change).[6] Upon completion of these tests the *S* was interviewed on the same attitude issue and also on his memory for, and feelings about, the hypnotic session. Finally the amnesia-removing cue was presented and after the *S* had achieved recall for the suggestion, the nature of the experiment was explained to him.

The full results of this experiment will be reported elsewhere.[7] It will suffice here to report that the original prediction was fully borne out and that the experimental *S*s showed significantly more cognitive reorganization ($p < .005$) than a control group not exposed to hypnotic suggestion.

Although it was not possible to analyze all of the resulting data in terms of the personality dynamics of the *S*s, with four of the *S*s circumstances were such that a large number of additional interviews could be conducted along psychiatric lines. With each *S* these interviews were focused on his conflicts, mechanisms, and personality-structure generally and also on his emotional reactions in, and defensive reactions to, the hypnotic process. Only when the dynamic interpretations drawn from these interviews were completed were they brought to bear upon analyzing the *S*'s performance in response to the specific posthypnotic suggestion he had received. In each case, the data on the *S*'s regnant conflicts and defense mechanism priorities and their influence upon his experience of hypnosis helped to illuminate the dynamics of his acceptance of, and adaptation to, the posthypnotic suggestion; and in each of these cases the resulting findings seem clearly to exemplify and confirm the hypotheses on posthypnotic suggestion that have been reported above. The next section of this paper is devoted to the presentation of these four cases.[8]

---

[5] All tests were administered by a person other than the *S*'s hypnotist.

[6] The full "cognitive structure test" records for all of the *S*s are reported in an addendum that has been deposited with the American Documentation Institute. Order Document No. 5721, from ADI Auxiliary Publications Projects, Photo Duplication Service, Library of Congress, Washington 25, D. C., remitting \$1.25 for 35 mm. microfilm or \$1.25 for 6 by 8 in. photocopies.

[7] In a forthcoming publication by one of the authors (M. J. Rosenberg).

SUBJECT D

Subject D was a young single woman, living alone in the city, and temporarily unemployed. She had been born and raised in Europe and at the time of the experiment had been in the United States for about five years. While she had an educational background and skills that suited her for an independent career, she had been employed in a field remote from her own. Relatively isolated, she was somewhat unapproachable, and heightened this defense by keeping as a pet an intimidating-looking dog. It was upon this dog that most of her affection was spent, and arrangements for her hypnotic sessions sometimes took second place to her providing care for her pet. She appeared to be in a chronically dejected mood, feeling sad, lonely, and suffering feelings of discouragement and emptiness. She had characteristically a wistful, pathetic manner, and elicited from those around her reactions of sympathy and gentleness, along with an awareness that she was still aloof and quite uncommunicative about herself.

The major historical items in her life are not essential here, but it is of interest to note that she came for hypnosis at a crucial juncture, and used it in a specific way. It was while working as a hypnotic *S* that she applied for, and got, a job somewhat closer to her interests; and it was during the course of the experiment that she arranged to do work, on a commission, for which she was really trained and which promised some success. Once this commissioned work was undertaken, she found herself unable to continue as a hypnotic *S* in a later experiment, and appeared to have made a more stable and more satisfying adjustment to her needs.

Among her original expectations about hypnosis was the notion that it would provide a relaxed, restful oblivion; something she yearned for but felt she had never been able to achieve. At the first hypnotic attempt, she went readily and "deeply" into hypnosis, exhibiting utter docility, and appearing to be a "perfect subject." This performance was not only gratifying to the hypnotist, but left the *S* with a feeling of profound exhilaration and satisfaction. Throughout the experiment there was a marked contrast between her rather apathetic, tired appearance before a hypnotic session and her comfortable, pleased look afterwards. She became enthusiastic about the experiment and practiced hypnosis upon herself at home. Throughout our work with her she would happily and spontaneously report upon her progress in inducing various hypnotic phenomena in herself. She understood her attempts at self-hypnosis as a kind of psychotherapy, a search for which was a subsidiary motive in her volunteering, in addition to her basic need for satisfying human contact.

Corresponding to her insulated mode of existence, the *S* displayed no strong sentiments on current or basic human issues. Except for her interest in her pet, which was protective and coddling, and a not always satisfying relationship to her work, she claimed no strong fixed opinions or attitudes.

Course of hypnosis. The major maneuver used by this *S* to surmount the dangers of hypnosis was the turning of a passively-experienced process into one actively reproduced by herself. That is, she reported consistently that it was not the hypnotist who gave her instructions under hypnosis, but herself. She exaggerated this not uncommon maneuver and carried it over into "waking" life, acting the part of the hypnotist to herself.

Although largely inarticulate about herself, she used the currency of her mood to communicate with the hypnotist. Thus her painful sadness at the outset, and her hypomanic state at the conclusion of a hypnotic session functioned as a plea and a reward to him. The appearance of submissiveness was enhanced by her frequent apologies for not "doing well enough," her fear that the experimenters would not be entirely pleased, and her eager promise to try to "do better" next time. In summary, she went to an extreme in almost abject compliance (against whose dangers she guarded herself by use of the mechanism described above) in exchange for the sensation of a consoling relationship with a powerful figure.

Adaptation to the posthypnotic suggestion. In this case, the posthypnotic suggestion was concerned with the *S*'s attitude toward the "city-manager plan." Prior to the experimental session she had indicated, on the attitude questionnaire, a moderately opposed position. Her score on the test of cognitive structure was −118, representing the fact that she perceived many bad consequences to follow from the city-manager plan.

---

[8] For each of the *S*s, the procedures that have been described were repeated for a second attitude area not dealt with in this paper. We were able to apply our interpretive procedures to the second attitude change for three of the four *S*s. In each of these cases the same individual dynamic patterns appeared as described here. With the fourth *S*, the change achieved was of too small a magnitude to enable analysis in depth.

In the experimental hypnotic session she was given the following suggestion of affect-reversal:

> When you awake you will feel very much in favor of the city-manager plan. The mere idea of the city-manager plan makes you feel very approving and enthusiastic. Although you will not remember this suggestion having been made, it will strongly influence your feelings after you have awakened.

Immediately following the trance in which this suggestion was delivered, the *S* was retested on the attitude questionnarie and on the test of cognitive structure. On the attitude questionnaire she now indicated that she was extremely "in favor" of the city-manager plan. This was accompanied by a spectacular shift on the test for cognitive structure to a score of +311. The *S*'s amnesia for the suggestion of affect-change was equally complete and thorough-going.

Her comments on the change suggested that the same mechanisms were operating as she employed in her hypnotic trance, along the lines of our Hypothesis I. Thus, in the *S*'s view, the change was accomplished by a "redefinition" of just what a city-manager plan was. She reported that prior to the experimental session, she had assumed that "it was a thing that restricted you—more or less enforcing all sorts of things." But, as she explained it, after the session, "I think I changed my assumption a little bit—it seems as though it all kind of reversed—I think really I feel the same way about it—only I just have stressed the ... another angle of it." Now, after the experimental session, the city-manager plan was perceived as "something voluntary which the people themselves would want to do for the city—and have everybody represented. I think maybe the first time [before the experimental hypnotic session] I thought the people wouldn't be represented and it would just be kind of inflicted on them—city bosses would get together and make decisions and possibly decisions that affect their personal interest instead of the people's welfare."

Later in the same interview she said, "at first I thought it would be a type of control which I have seen which I heartily dislike—for instance in cities where they invite all sorts of industries to come regardless of what type it is. And then the second time I had in mind a sort of more democratic system in which people take into account each others interests." She reported having been unaware of the "democratic" definition of the city-manager plan prior to the posthypnotic suggestion. She attempted to explain this change in the meaning of "city-manager plan" on the basis of her having "whims," being changeable, and said of herself, "I'm not terribly consistent in the way I think."

The attitude-change suggestion, then, has been complied with by treating it precisely as she has treated the hypnotist's activity in hypnosis: by constructing a "redefinition" of the concept "city-manager plan" whereby it loses its authoritarian, mastering implications, and acquires a measure of voluntarism. This more generous view of the city-manager plan as "democratic" corresponds to her more comfortable acceptance of hypnosis when she can turn a passively experienced event into one actively reproduced by herself so that, in effect, her interests are observed and taken into account. This shift in the meaning of "city-manager plan" can also be seen as a carry-over of her enduring affective state in hypnosis; a state of a pleased and grateful feeling toward the hypnotist who has become, not a tyrant, but a kindly figure. In brief, the adaptation to the posthypnotic suggestion seems, in this *S*, to be a rather clear illustration of the operation of the dynamic process which is posited in Hypothesis I above.

Her changes on the test of cognitive structure add confirming detail to the interpretation that has just been given. For example, on the test before the experimental hypnotic session she judged the value item, "complying with the wishes of persons in authority," as being "dissatisfying" (rating of −8 on the value satisfaction dimension) and indicated that she believed the city-manager plan tended to foster this negative value. After the administration of the posthypnotic suggestion, this same value item, while still judged as one which is dissatisfying, is now seen by the *S* to be "blocked" and prevented by the city-manager plan. For this *S*, the value item of "complying with the wishes of persons in authority" is probably representative of the idea of loss of autonomy. By transforming the meaning of the city-manager plan into a concept stressing its "democratic" features, she is able to deny the element of compliance both with regard to the posthypnotic suggestion and with regard to hypnosis itself.

A similar change occurred with the value item, "having power and authority over people," which is considered by the *S* to be a dissatisfying thing and is seen as attained through the city-manager plan before the posthypnotic suggestion and as blocked by it after the posthypnotic suggestion.

The mechanism she has used—of becoming her own hypnotist and thus effecting a successful merging of a benevolent, sleep-giving image of the hypnotist with the self—is reflected in four value items which, prior to the trance, were seen as blocked by the city-manager plan but after the trance are seen as fostered and attained by the city-manager plan. These four value items are "all human beings having equal rights"; "the open expression of disagreement between people"; "people of different backgrounds getting to know each other better"; "people having the right to participate in making decisions which will affect them." Her posthypnotic classification of these four values seems to bring to bear upon the task of rationalizing the posthypnotic suggestion, the rewarding closeness and equality that she has achieved in hypnosis through her mechanism of denying that hypnosis involves any domination.

In these and other aspects of her rationalization of the affective posthypnotic suggestion, the *S* gives further evidence that for her, acceptance of the posthypnotic suggestion is guided by the type of dynamic process that we have pointed to in Hypothesis I: the carry-over of the same mechanism to the posthypnotic suggestion as operates in defending the *S* against the danger points of hypnosis.

In view of the lack of substance of her fixed attitudes, Hypothesis II is not called into play. As she herself states, she generally has little investment in social questions and in fact may adopt or relinquish a point of view in order to please or conform with an expectation of someone else. Thus she appeared interested in the details of her attitude change, but did not feel strongly about the fact that she had changed. Rather she was more concerned about continuing work as a hypnotic *S* in a further experiment. Her greatest satisfaction after the experimental session appeared to be that she had performed well and had pleased the hypnotist.

SUBJECT F

Subject F was a 21-year-old single student who was in his last year of preparation for divinity school. Of the four *S*s whose hypnotic performances are analyzed in this paper, he is the one on whom we have the least amount of clinical material. But interviewing and observation were sufficient to suggest certain salient facts about his character structure and his techniques of defense. The overall picture seemed to be of a sado-masochistic person in whom the instinctual forces of destructiveness were tightly bound and rechannelled through a life pattern based on reaction-formation. While acknowledging that he had no true "religious faith" and felt considerable skepticism about the dogmas of his and other denominations, he saw the ministry as a type of work through which social good might be done. For him its ethical imperatives were the source of such validity as Christianity possessed.

During his years as a student he had held a part-time job as an attendant in an inpatient psychiatric hospital and found this work "fascinating." It would appear that in general he was able, through the mechanisms of denial and reaction-formation, to control his sadistic impulses by transforming them into an ideologized charitableness and "good will." These he expressed in part through his involvement in the functioning of the psychiatric institution; but in this setting his superordinate relationships with the emotionally ill allowed him, at the same time, a continued gratification (through the secret pleasures of "Schadenfreude") of the sadistic component.

His style of conversation, if expressive cues were ignored, was friendly and cooperative, sometimes diffident. But by the use of his voice and his tendency to produce long mid-sentence pauses during which he would scowl, examine the room and then, in muffled tones, continue his comment, he seemed to reveal a persisting suspiciousness and an emotional substratum of irritability.

From his first contact with the experimenter he expressed great interest in hypnosis and an eagerness to "see for myself what the hypnotic experience is like." He was able from the first session to accept a degree of hypnotic involvement that included the achievement of motor effects, anesthesia, compliance with posthypnotic suggestions and posthypnotic amnesia.

Course of hypnosis. During each of the training sessions, the *S* entered rapidly into the trance state. A common accompaniment of entry into trance was loud deep breathing and a heavy thrashing about of limbs and trunk. The posthypnotic suggestions were almost always carried out with a rather exaggerated compliance to the implications as well as to the letter of the instructions. Thus when told that a certain book would appear to be printed in German he shouted upon first opening its pages, "This is fantastic—it's in German." He then proceeded to read aloud a paragraph from this book with the experimenter looking on. His "reading" consisted of a rapid and fairly accurate sight translation, into German, of the English text.

In the discussions that followed each training session he would first manifest a sense of dullness and confusion, sometimes seeming rather stuporous. On one such occasion he commented that he was not sure whether he was or was not still in hypnosis. As his confusion wore off he would begin to raise questions about the nature of hypnosis. In doing this he seemed to be indirectly expressing and stilling his anxiety about the intense relationship into which he had allowed himself to enter.

The contents of the posthypnotic conversations and of the later interviews suggest strongly that for this *S* hypnosis was a very fulfilling sadomasochistic encounter; one in which by total "submission" he achieved a rewarding identification with the hypnotist and thus partook of the hypnotist's magical powers of dominance, control, and omnipotence. His pressing needs for such sadomasochistic fulfillment were, however, almost totally inadmissable in terms of his conscious self-image and his usually effective employment of the mechanisms of denial and reaction-formation. The resulting, and potentially disruptive, conflict he usually handled by the expedient of dullness and inarticulateness in the period immediately following emergence from the trance state. In this way he was able to isolate the ego state obtaining during hypnosis and the motivational forces with which it was identified as though they referred to a different person; i.e., an amalgam of self and hypnotist. He thus avoided the requirement of integrating these unassimilable impulses with his usual self-image.

Adaptation to the posthypnotic suggestion. With this *S*, the posthypnotic suggestion of affect-reversal was focused on his attitude toward "Negroes moving into white neighborhoods." Prior to the experimental session he had indicated on the attitude questionnaire instrument a strong positive reaction to "Negroes being allowed to move into white neighborhoods." However, his total score on the test of his cognitive structure about the housing desegregation issue was +3. This score was based upon his holding separate beliefs that Negroes moving into white neighborhoods would have both positive and negative consequences. The discrepancy between his strong positive score on the test of attitudinal affect and his mixed and ambivalently organized cognitive structure suggests that he felt required (whether by the situation or by his own ego ideals) to represent himself as fully committed to the prodesegregation position, whereas he was in fact capable of thinking about the issue in a way which reflected some antagonism to the possibility of Negroes moving into white neighborhoods.

In the experimental hypnotic session, he was given the following suggestion of affect-reversal:

> When you awake you will feel great loathing and disgust toward the idea of Negroes being allowed to move into white neighborhoods. The mere idea of Negroes being allowed to move into white neighborhoods offends you and makes you feel quite angry. Although you will not remember this suggestion having been made it will strongly influence your feelings after you have awakened.

While listening to the instructions about his posthypnotic feeling on the issue of Negroes moving into white neighborhoods, the *S* emitted a series of rasping, growl-like sounds. These ceased after the presentation of the suggestion was completed. Immediately after emergence from the trance and before the readministration of the tests could begin, the *S* began to manifest various expressive signs of anger. He kept his lips tightly pressed together, his fists clenched and spoke with a clipped, stacatto rhythm.

On the second administration of the attitude questionnaire, he marked as his own the extreme negative position with regard to Negroes moving into white neighborhoods. In filling out the Negro desegregation item he continued to show high tension, held the pencil very firmly, rewrote his answer several times and then placed a number of asterisks next to it. As the testing continued, he frequently broke into spontaneous comments of hostility toward Negroes, either shouting or snarling most of these comments. For example, at one point during the cognitive structure test he said with considerable fury, "I'm prejudiced. I'm against Negroes moving into white neighborhoods. They should stay in their own dirty, sloppy rundown neighborhoods and not spread destruction into white neighborhoods." In the interview following the completion of the test procedures, the *S* continued to express hostility and condemnation toward Negroes moving into white neighborhoods.

Despite his frequent outbursts against Negroes, his behavior during the cognitive structure test was confused and disordered. It was as if, although expressing very strong anti-Negro affect, he was incapable of rationalizing that affect through the provision of arguments that justified it. The card-sorting test of cognitive structure, which he had taken and fully understood before the experimental hypnotic session, now seemed to confuse and overwhelm him. He completed it only by throwing the cards into categories in an almost haphazard fashion, all the while voicing the complaint that the test was confusing, senseless, and boring. This general avoidance of the rationalization task posed by the test of cognitive structure renders his specific card-sorting performance irrelevant for the present analysis. But the *fact* of this avoidance is of great interest in interpreting the *S*'s general adaptation to the posthypnotic suggestion.

It has already been suggested that this *S*'s easy entry into, and tolerance for, hypnosis is based upon the opportunity for gratification of his sadistic wishes. Through his identification with the hypnotist he is experiencing in hypnosis that sense of manipulative and potentially injurious sadistic control which he ordinarily transforms through denial and reaction-formation. Now in the experimental session he is told by the hypnotist that he will feel negatively toward Negroes moving into white neighborhoods. This suggestion gives him permission to enter into a virtually uncontrolled outburst of primitive rage, and to direct that rage at one of those objects which, by virtue of social tradition, has always been available for attack if the attacker can tolerate acting out his own hostility. Such a tolerance for his own sadistic impulses is not normally possible for this *S*, but because of the release that hypnosis has already provided and because of his temporary assimilation of an image of a strong, sadistic, and responsible hypnotist, he is now free to enjoy, in comparative safety, the release of his sadistic impulses.[9]

Thus the dynamic process that seems to govern this *S*'s acceptance and execution of the posthypnotic suggestion of antidesegregation is the one encompassed in our second hypothesis. The posthypnotic suggestion is effective through its affording a comparatively safe abandonment of a basic reaction-formation; an undoing and temporary cancellation of ego control which frees, for direct expression, a driving force of primary significance in the *S*'s emotional life.

The near violence with which this *S* expressed the anti-Negro-desegregation affect is probably of significance in accounting for his inability to rationalize this affective change through his responses on the cognitive structure test. We have seen that the *S* maintains a posthypnotic pattern of dullness and confusion that enables him to keep the satisfactions of hypnosis discrete from his usual picture of himself. Similarly, the strength of his reaction-formation to his sadistic impulses requires that he keep the outburst of anti-Negro feeling out of connection with his usual ego state. Using the same mechanism of functional stupidity, he finds the cognitive-structure test confusing and himself dull and muddled. Thus he avoids having to offer an argumentative defense of the anti-Negro affect and this avoidance enables him to keep his expression of that affect out of association with his usual patterns of thought. In this aspect of his adaptation to the posthypnotic suggestion we seem to have the sort of dynamic event dealt with in Hypothesis I, namely: an execution of the posthypnotic suggestion through the employment of the kind of mechanism which is regularly used by the *S* to overcome the danger-points of hypnosis and to maintain himself in the hypnotic relationship.

After the testing was completed, the experimenter had a further discussion with the *S* before giving the amnesia-removing cue. Despite numerous questions about his reasons for feeling so strongly against Negroes moving into white neighborhoods, the *S* could not produce any coherent rationalizations except in one highly interesting exchange which is given here:

> Hypnotist: What do you have against Negroes moving into white neighborhoods?
>
> Subject: What do I have against it?
>
> H.: Yes.
>
> *S.*: Oh for heavens sake. The evaluation of property goes down, education is . . . is . . . and look what happens to our churches, they fall apart.
>
> H.: Do they really? It's something I don't know much about.
>
> *S.*: Why today the downtown church is a ruined church because of the Negroes moving into city areas, the people who . . .
>
> H.: The downtown—you mean on the Green?
>
> *S.*: Those aren't typical churches there. They have sense enough not to let Negroes come to those churches.

What is of particular interest in this excerpt is that virtually the only rationalization that the *S* could muster is one that involves a reference to "the church"; the very institution which in his conscious thought is the source, and effective instrument, of those moral imperatives which bind and transform his sadistic tendencies. Possibly he is working off some guilt over his aggressive outburst; guilt which has not been fully and successfully blocked by his restriction of the anti-Negro affect to a somewhat altered state of consciousness. Thus, to combat this guilt he improvises a perception of Negroes as threatening the agency that symbolizes those very moral standards which he has violated by expressing his sadistic impulses.

When the amnesia-removing cue was given, the *S* had complete recall for the original suggestion, immediately spoke of feeling great embarassment, and then went on to comment at length on how interesting it was that hypnosis could make one behave in a way so unlike one's "real self." All indications were to the effect that his final resolution of the incident was to once again seal over the sadistic impulses that he had indulged, so that his defensive structure remained easily as intact as it had been before the experimental session.

SUBJECT J

The *S*, a married man with three children, was a conscientious, hard-working person of high intelligence, with a very good academic record behind him. He was attending professional school at an age somewhat older than usual, having had a career in a related field which did not satisfy him enough. He felt that his earlier area of work was only second-rate and, although worthwhile in itself, lacked the prestige of the profession he was now preparing to enter. His attendance at school had been at the cost of considerable sacrifice to himself and to his family; his wife had held two jobs at the same time to help support the family.

---

[9] It seems possible too, that the intensity of the *S*'s rage is due partially to a *displacement*, onto the content of the hypnotist's suggestion, of strong anger felt toward the hypnotist. Such anger, based upon the manipulative, intrusive aspect of the hypnotist, cannot be directly expressed if hypnosis is to be continued. But it can be displaced by virtue of easy associative paths, and by virtue of its congruence with underlying sadistic impulses, to the hypnotist's suggestion.

After a prolonged period of scraping by financially, Subject J was now able to proceed with further professional training because of outside help given him on the strength of his previous record. He felt himself to be on the verge of professional rise and fulfillment. He had aspired for many years to this kind of distinction that would set himself apart from and above his peers as an educated and competent person. He described himself as wanting "to be a big boy." Among the inner obstacles to his achieving this status were his awareness of some of his own dependent, "little boy" characteristics, and his considerable shame about his parents, their way of life, and his "messed up" home life as a child.

He had felt his family, in which he was the only child, to have been a source of shame and a secret to be kept from his friends. There were thoughts and ruminations about his family that he feared to reveal, feeling that were they known they would prejudice his chances of professional eminence and possibly ruin him or expose him to ridicule. In this connection he once said that now that both his parents were dead, it did him no good to talk about them since it would only gain him sympathy when what he wanted was advancement. He had sought this advancement steadily and successfully, but at this moment in his life found himself in the paradoxical situation of having professional success on the way and yet being dependent on financial help from others after having fought most of the way alone and unaided. He was proud of getting so far, of having pushed himself to his present level of competence and recognition.

However, his present family responsibilities seemed somewhat of a burden, and he complained at times of his irritability and curtness toward his children. There had been more serious sources for this irritability in the past; at one time, one of his children had required major surgery while his wife was ill and incapacitated. At the same time, his father, whom he had privately renounced and had always ambivalently loved, was in a terminal illness, and the *S* was required to bear part of the financial burden for that as well.

The *S* volunteered for hypnosis in response to a current need: to find a stronger and more respectable figure than his father with whom to participate and with whom to share, especially through work with professional and academic overtones. But at the same time, this father-substitute had to be debased and scorned so as to enable the *S* to assert his own independence and power. Illustrative of this conflict was the fact that very early in the experiment he endeavored to discover what degree the hypnotist held and was gratified when he found that, in his eyes, it did not carry as much prestige as the degree he himself was soon to receive. This ambivalent regard for the hypnotist was matched by some vacillation over attaching himself to him. He "confessed" during a later conversation that in preparation for one of the hypnotic sessions he had indulged in one of the habits that he detested in others: he had "braced" himself with a number of cocktails.

His characteristic way of dealing with affects was to avoid expressing them except in detached, intellectualized descriptions. Thus, he could describe suffering "dreadful fear," and itemize this with an accurate account of associated somatic manifestations while yet remaining visibly composed and even intrigued by his description. In most affairs he adopted the basic stance of a detached, almost-clinical observer, setting himself above the emotions he might feel. He had an eager readiness to introspect, to think of himself in psychological terms, to outdo the hypnotist's expectations in his candidness and open-mindedness about his inner life. Later interviews with the *S* confirmed the hypnotist's impression that by the clinical stance that he took toward himself, the *S* was attempting to reward the hypnotist and thus assure a prolongation of the relationship. At the same time, this aspect of the *S*'s behavior is a good example, of which there were many others, of his assuming some traits of the hypnotist by way of identification.

Course of hypnosis. The *S*'s fears about hypnosis concerned its possibly "spectacular" influence, its being somehow faintly improper, and its being a form of probing and "brain-picking" through which something about him might be revealed and even publicized. This last mentioned fear stood in distinction to his usual eagerness about his own "inner life." In addition to these there were fears corollary to his basic motivation for hypnosis: his need for association with a "big boy" of some prestige whom he can yet master; with whom he can be close, but still safe. These were fears of being dominated, overpowered and overcome.

The "danger points" in hypnosis were easily demarcated by the *S*. The first step in induction seemed frightening, even alarming at times, by virtue of its being a signal of his submissiveness. This experience of fear was often colored with ideas of too great intimacy, too sensual or sexual a closeness, and what he called a "threat to my masculine feeling." These fears (at one time he described hypnosis as being "like witchcraft, as though it smacks of evil") were dealt with by heterosexual fantasies of "sweater girls," and by his usual mechanism of robbing the situation of affect and converting it into something he himself was observing in an aloof way. The tests of hypnotic performance, eye closure for instance, were equally alarming to him, as were his experiences of "falling" and other changes in body feeling. At these times he would be aware of a lightening of his trance. Another way of dealing with these and similar danger situations was through identification with the hypnotist whom he sometimes saw as a protecting, feeding person. In his identification he *became* the hypnotist, lost any feeling of the hypnotist's nearness, retained only his words but not his voice as such. According to his report, it was "as though I thought the suggestions myself." Nevertheless, he often got the words wrong, and would occasionally carry out a suggestion much distorted, even caricatured or reversed. In this process he seemed to be trying to deny and limit the influence of the hypnotist. Despite this, he was an unusually adept *S* in most respects; the readiness with which he could be hypnotized eventually became his chief source of great anxiety.

His general attitude toward the experience of hypnotic trance is partially summarized in imagery he produced during one session in which he visualized a little boy being given bread and patted on the head and then assaulted cruelly by a "big monster." The fantasy ended with the death of the little boy under the wheels of a truck.

To recapitulate our analysis of this *S*'s adaptation to hypnosis: he seeks a father-replacement or substitute at a time when his usual reliance on his own great efforts for advancement is suspended by his receiving outside help. Distrustful of the intrinsic strength of his father-substitutes, he attempts to hold the hypnotist to him while at the same time belittling him; he scorns the hypnotist while at the same time making an identification with him. The dangers of the passive-dependent role in hypnosis are dealt with by such methods as avoidance of affect, the cultivation of a detached "observer" role, isolation, and a reassertion of his own personal strength.

Adaptation to the posthypnotic suggestion. The posthypnotic suggestion used with this *S* dealt with his attitude toward "the federal government providing comprehensive medical insurance for all persons." Before the experimental session, the *S* had characterized himself, on the attitude questionnaire, as being extremely in favor of federal medical insurance. His score on the test of cognitive structure was +195, representing the fact that almost all of his stated beliefs about federal medical insurance were of the positive variety.

In the experimental hypnotic session, the *S* was given the following suggestion of affect-reversal:

> When you awake you will feel great loathing and disgust toward the idea of the federal government providing comprehensive medical insurance for all persons. The very thought of this plan offends you and makes you angry. Although you will not remember this suggestion having been made, it will strongly influence your feelings after you have awakened.

Immediately following the hypnotic session the *S* was retested on the attitude questionnaire and the test of cognitive structure. As measured by the attitude questionnaire, the *S*'s attitude toward federal medical insurance was now given as *mildly* positive (this change representing a significant shift in the negative direction). A more extreme change was noted in the test of cognitive structure. From +195 before the posthypnotic suggestion, he moved to −58 after the posthypnotic suggestion, a score based on a cognitive structure containing beliefs of both positive and negative consequences following from federal medical insurance. In general, the posthypnotic suggestion in this case had the effect of rendering *ambivalent* an attitude that had previously been purely positive.

Immediately following the second testing, the hypnotist engaged the *S* in a conversation dealing with his feelings about federal medical insurance and about the experimental hypnotic session. Two particular themes recurred frequently during this conversation. The first of these themes was revealed in the *S*'s numerous references to the fact that it was only "the second time around" (i.e., after the experimental hypnotic session) that he came to realize that federal medical insurance was not a pure blessing but, rather, a mixed thing. The negative aspect of federal medical insurance which he now saw clearly and felt concerned about was that "it could infringe upon the rights and liberties of individuals." Frequent references to this idea marked it as the major area of change in the *S*'s posthypnotic thinking on the issue of federal medical insurance. It is of course noteworthy that in his earlier adaptation to hypnosis, the *S*'s major problem was to overcome his sense of danger at giving up the independence, autonomy and emotional distance that were so crucial to his general adjustment.

An interpretation that seems to us to be required in this case is that the *S* manages partially to interiorize the suggestion of negative affect toward federal medical insurance by finding that this object violates those very standards (of independence and autonomy) which he feels to be threatened by hypnosis itself. Thus in partially accepting the posthypnotic suggestion he is at the same time defending himself from the danger of loss-of-autonomy in hypnosis by rationalizing the suggested affect in terms of those standards which figure in his characteristic stance towards hypnosis. The process involved, then, would seem to be the one identified in Hypothesis I above.

A second theme that arose frequently in the discussion after the experimental hypnotic session was the *S*'s awareness of "feeling angry." As he reported it, he had awakened from hypnosis "feeling a sort of clenched fist kind of anger." He remarked that perhaps this anger had been "set up" for him by the experimenter. Later when the amnesia was lifted, he felt the same sort of anger surge back when the experiment was explained to him. But though he described himself as feeling angry, he did not allow himself to express this feeling except by intellectually noting its presence as a subjective state.

A further complication in expressing his sense of anger at having been manipulated is revealed in the fact that, immediately after having mentioned feeling angry, he went on to express his "utter amazement" at the fact that his opinions had been changed by the hypnotist. This comment was followed by a repetition three times of the phrase, "Isn't that wonderful?"

What the *S* seems to be doing in this interlude after the lifting of the amnesia is to give continued expression of his mixed reaction to the hypnotist by way of commenting on the hypnotist's experiment. It seems to us possible that the *S*'s ambivalence toward the hypnotist and toward the hypnotic experience is reflected in his responding to the posthypnotic suggestion by developing an ambivalent attitude towards federal medical insurance. In other words, the content of the hypnotist's suggestion has been treated in the same way as the figure of the hypnotist; the content of the manipulation has been dealt with in the same way as the fact of manipulation. Such a process could be understood in terms of our Hypothesis I.

Both aspects of our interpretation seem to be borne out by a number of changes on the test of cognitive structure. For example, on the first test (before the experimental hypnotic session) he judged the value item, "people having the right to participate in making decisions which will affect them," as being neither satisfying nor dissatisfying to him (i.e., a rating of 0). He indicated further that, as he saw it, the institution of federal medical insurance would tend to lead to the attainment of this value. But after the administration of the posthypnotic suggestion of negative affect toward federal medical insurance, two changes are noted: the value increases in its satisfaction rating from 0 to +8, demonstrating what he says about the increased importance in his thinking of matters of individual rights; while formerly seen as attained through federal medical insurance, this value is now seen as blocked and prevented because of federal medical insurance. With the changes in his thinking about this particular value, the *S* is apparently rationalizing the posthypnotic suggestion by a reassertion of those very standards of autonomy and equality that he regularly invokes against the danger of giving way in hypnosis to a full indulgence of his need for submission to a "big boy." Viewed in terms of Hypothesis I, we have here the transfer of an already established defensive operation onto the new intrusion of the hypnotist; that is, onto the posthypnotic suggestion.

A sampling of some other changes in cognitive structure underscores the parallelism between the manner in which he accomplishes his negative shift toward federal medical insurance and the manner in which he succeeds in hypnosis. Thus the value item, "all human beings having equal rights," which he judges as being attained through federal medical insurance before the posthypnotic suggestion, increases in its satisfaction rating from 0 to +9; but after the posthypnotic suggestion, this same value is seen as blocked by federal medical insurance. This change may be viewed as reflecting his sense of threat at the domination implicit in the hypnotic relationship and in the posthypnotic suggestion. It is through the process referred to in Hypothesis I that he manages, at least partially, to accept and execute the posthypnotic suggestion of negative affect toward federal medical insurance by seeing this social institution as a threat (just as are hypnosis and the hypnotist) to the rights of individuals.

This process of transforming the meaning of federal medical insurance to something akin to his picture of hypnosis is seen also in the following cognitive change: the value item, "having power and authority over people," is for him a negative (dissatisfying) value both before and after the posthypnotic suggestion; but while it is judged, before the experimental hypnotic session, as being blocked and prevented by federal medical insurance, after the posthypnotic suggestion it is judged as attained and facilitated through the effects of federal medical insurance. Once again a change in a portion of the *S*'s cognitive structure reflects the fact that his rationalization of the posthypnotic suggestion is made possible by transferring to the concept of federal medical insurance the pattern and content of his ambivalence toward hypnosis itself.

The *S*'s resentment at hypnosis and his preference for an independent course for himself are reflected further in his judgments of the value item, "the open expression of disagreement between people." This item is judged, both before and after the experimental hypnotic session, as being blocked and prevented by federal medical insurance. However it changes in its satisfaction rating from 0 to +9 and so helps turn his attitude against federal medical insurance.

In general, the changes in the *S*'s beliefs about federal medical insurance are shifts reflecting his own emphasis on individual initiative, strength, and autonomy. In complying with the posthypnotic suggestion he experiences federal medical insurance as something threatening these values, just as for him hypnosis threatens them. Nevertheless, his affect and his beliefs with regard to federal medical insurance remain mixed and ambivalent just as do his affect and beliefs pertaining to hypnosis. The over-all process, then, by which this *S* seems to have accepted and worked through the content of the posthypnotic suggestion seems clearly to exemplify the dynamic operations dealt with in Hypothesis I.

The kind of process referred to in Hypothesis II does not seem to operate in this case. This is probably because it is the *fact* of manipulation to which the *S* responds rather than to the content of the specific attitude being manipulated.

SUBJECT R

Subject R was a young single graduate student, chronically depressed in mood but not given to overt complaining about it. He was almost obsessively aware of his relatively humble origins in a small midwestern town and of his disappointment in his family who did not fully understand or sympathize with the education and career aims he had set for himself. He felt estranged from and out of contact with his life at home. He was devoted to study and to proving himself intellectually, and in these activities he had been quite successful. Socially, he had always been isolated and inept.

He came as a volunteer *S* for hypnosis at a time when he was searching urgently for a satisfying relationship with a girl but feeling disappointed and enraged at his usual frustration and rejection at the hands of women. He had customarily antagonized or at least frightened off potential friends, both male and female, by adopting a protective appearance of self-satisfaction and conceit or else by unburdening himself of long and tortured accounts of his difficult childhood. Such accounts did, at times, evoke sympathy from others but were ordinarily taken as indicative of his having a reproachful, gloomy nature.

He was envious of his school friends who were sometimes boisterous and often boastful of their sexual exploits; but at the same time he condemned any kind of license, laying stress on propriety and very rigid moral standards. At the time he volunteered, he was suffering a period of distressing sexual fantasies. These fantasies were of a wide variety and were experienced especially at night, around bedtime. He attempted to cancel these out by deliberately preoccupying himself with ideas of clean, wholesome, "pure," and largely asexual affairs with girls. During the day these preoccupations were put into the background by his intense studying, which he sometimes used deliberately for this purpose, although he could not devote his full attention to his work. He *was* able to augment his self-esteem by his frequent academic achievements, and he clung eagerly to these as sources of great satisfaction.

Extremely sensitive to any hint of rejection, the *S* dealt with the hypnotist much as he did with others in his life: by constantly increasing demands for time and understanding, and by expressing rage when slighted. It was clear that one of his underlying motivations for hypnosis was the hope that through its therapeutic effects, he might eventually be enabled to establish the kind of "pure" heterosexual relationship he yearned for. This yearning was so encumbered with painful regressive fantasies, and its attainment seemed so prohibited to him, that he settled into a relationship with the male hypnotist as an alternative. It was striking to note that he was the only *S* in a large number who not only made note of the hypnotist's first name, but used it exclusively, implying not only friendship but also a social-age equality.

In retrospect, his initial reaction to hypnosis was understandable: he "forgot" the first appointment, having made a date on short notice with a girl for that same time. This date, as it turned out, was as unsatisfying for him as any other and, in fact, during it he conducted himself in a somewhat bizarre and conspicuous manner.

It is of particular pertinence to our analysis of this *S*'s hypnotic performance that he described himself as disgusted at the licentiousness, filth, and sensuality of city life, and at the sexual seductiveness of girls. He contrasted this with the cleanliness, simplicity, and piety of his rural childhood home, and ultimately with his picture of a pure and untroubled childhood. It was clear that his impression of the city was largely a product of the projection of his own sexual fantasies which, on examination, were found to contain ideas of biting, injury, anal, and urethral soiling. Such fantasies and projections also figured as part of the background for his hostile or diffident wariness with girls. Following his work as a hypnotic *S*, and perhaps affected by the considerable amount of therapeutic handling of his feelings in the psychiatric interviews, he established his first enduring relationship with a young woman. But on entering hypnosis, the central issue was his loneliness, mediated by the unconscious elements described above, and his shrinking from any mature heterosexual interests or society.

Course of hypnosis. The entire period of induction was full of danger points for this *S*. He proceeded slowly and painstakingly through each step; eye closure, tests for "depth" of hypnosis, and the trance. The fear of being under someone's spell, as fixed by a memory of a movie about hypnosis, was the initial danger. At a deeper level he felt considerable terror at the prospect of "loosing control" over his impulses. These dangers he tried to cope with by his deciding, during induction, and in a deliberate fashion, that his motivation for hypnosis outweighed his fear. However, at the first evidence in himself of compliance, he was overcome with distressing somatic sensations, all of which were connected in his mind with sexual stimulation. The most constant of these was a change in his sense of equilibrium. This was often accompanied by ideas of mutilation and assault. At other times, the danger presented itself in the form of spontaneous imagery expressed in anal and urinary terms.

His response and accommodation to these dangers was manifold: in a manner characteristic of his waking life, he would seek protection from the feared power of the hypnotist and from his own impulses by putting emphasis on cerebration and intellectual control. This was graphically demonstrated in his habit of focusing his "mind's eye" on a dark spot that he identified as his "frontal cortex" and that symbolized for him the control of unruly thoughts. At other times he sought such control by the pseudo-hallucinatory evocation of the figure of a cross. Typically the cross (which was a powerful link with his pious and innocent childhood) was carried over from one of the last things seen before closing his eyes; an intersection of joints on the ceiling.

At a deeper level, his fear of losing control was extended to a fear of "going crazy," and, it was discovered, this was for him an unverbalized daytime fear as well. To counter this fear he would sometimes manage a reassertion of contact with the reality about him. As in the case of the cross, he would often take with him into hypnosis some tag of the room, the day, or the hypnotist to cling to, and these served as reassurance of a stable and orderly world.

At all points in induction where he felt that he was "going deeper," he would actively fantasy himself in the role of the hypnotist. For example, he would imagine "someone" lifting his arm and would find that this changed imperceptibly to a picture of himself picking up his own arm. He responded slowly to any intervention or suggestion of the hypnotist, explaining this slowness of response, after hypnosis, as due to the mental effort involved in making the hypnotist's suggestions appear to be his own suggestions. On the hypnotist's speaking to him (he usually "heard" the hypnotist as speaking loudly, or even shouting) he would become angry, feel intruded upon, and then, after a process of inner debate in which he weighed the pros and cons, he would carry out the suggestion as though he had decided upon it spontaneously. The hypnotist's intrusion was dealt with, then, by an identification with the aggressive aspect of the hypnotist's behavior. Further evidence of this identification is seen in the range of imagic impressions of the hypnotist that he reported having had from time to time during trances: sometimes the hypnotist was imagined as a policeman who was not a punishing figure but rather an adjuring and protective authority; sometimes as a loving, fatherly person who was understanding and withheld punishment; and finally, as a feeding and nurturing, motherly figure. This last image merged and coincided with fantasies of eating, of breasts, and of landscapes reminiscent of his childhood. During the times when induction resembled his hypnogogic revery and recalled its accompanying sexual excitement, the hypnotist appeared as a too-permissive seducer who was failing to keep his sensuality within bounds.

The potential gratifications of hypnosis were hinted at in his frequent preoccupation with it as a kind of test or course in which he must do well, keep the favor of the hypnotist, and chalk up a success for himself. In this way he might achieve a restful, relaxed, and sleepy state near the end of his trance, pleasing both the hypnotist, with whom he was identified, and realizing his own hopes for a therapeutic stilling of his anxieties and depressive agitation.

In summary, the *S* reacted to hypnosis as a potentially rewarding kind of relationship which, despite its dangers, was marred only by the hypnotist's interventions; these were deeply experienced as a form of sexual arousal or assault. Such threats he could accommodate to by identifying himself with the intrusive aspect of the hypnotist, hoping thereby to preserve the hypnosis and the consoling advantages of his contact with the hypnotist.

Adaptation to the posthypnotic suggestion. With this *S*, the posthypnotic suggestion dealt with his attitude toward "living in the city of Los Angeles." Before the experimental session he had indicated on the attitude questionnaire that he was strongly opposed to living in Los Angeles. His score on the test of cognitive structure, however, was +47; in his thinking about Los Angeles, he associated it with a few more "good consequences" than "bad consequences."

In the experimental hypnotic session he was given the following suggestion of affect reversal:

> When you awake you will feel very much in favor of living in the city of Los Angeles. The mere idea of living in Los Angeles makes you feel happy and exhilarated. Although you will not remember this suggestion having been made, it will strongly influence your feelings after you have awakened.

In the retest immediately following the hypnotic session, the *S* now felt very much in favor of living in Los Angeles, choosing the most extreme pro-Los Angeles statement. The score based on the second test of cognitive structure was +137.

Before any attempt was made to remove the amnesia, the *S* was asked whether he remembered having been told anything about Los Angeles. He could not recall having heard any such comment but did go on to say that he had "always wanted to go to California." It was not until well after this session that it was learned (in a later interview) that Los Angeles had always occupied a special place in his thoughts as epitomizing license, promiscuity, and instinctual release of all kinds; that it was not an indifferent place-name to him but evocative of the same impulses which he strove so hard to control in himself. In this connection it is striking to note that, during the experimental session, even after the amnesia-lifting cue had been given, he did not remember the posthypnotic suggestion having been made. He continued to muse that probably Los Angeles is a very nice place to live and said, "Well, if anyone told me I had to go there, I'd probably go gladly."

The posthypnotic suggestion used with this *S* would appear, then, to have touched upon an issue derived from one of his major reaction-formations. His easy compliance with the suggestion, despite the fact that he was frequently a difficult *S*, would appear to be due to its having provided him, within the context of the special meanings of the hypnotic situation, with a disguised and gratifying way of expressing strongly conflicted sexual impulses; there were sexual impulses with which he was currently struggling. Thus, the *S*'s easy and rather thoroughgoing compliance with the posthypnotic suggestion is to be understood in terms of the process dealt with in Hypothesis II above.

Further evidence in support of this interpretation is to be found in the data gathered through the two administrations of the test of cognitive structure. For example, on the test before the experimental hypnotic session the *S* had judged the item, "people having strict moral standards" (for him a "positive" value), as being blocked through living in the city of Los Angeles. But after the experimental session in which, through the posthypnotic suggestion, he has been given permission to abandon this defense, this value item is no longer seen as affected by living in Los Angeles; i.e., the belief that living in Los Angeles lowers one's moral standards no longer figures as a reason for disliking Los Angeles. Thus, in complying with the posthypnotic suggestion, he has not only identified himself with what seems like a permissive and libertarian attitude in the hypnotist, but has relinquished some of the defensiveness that he had previously centered around this issue.

An equally meaningful set of changes is observed in the case of three value items, all of which, on *both* administrations of the cognitive-structure test, were seen by the *S* as being *attained* through living in Los Angeles. In the case of these items, the significant change lies in the fact that the satisfaction rating of each of them went up by five or six points from prehypnotic to posthypnotic administration. These value items were: "change and variety: having new kinds of experience," "being with other people; socializing," "people of different backgrounds getting to know each other better." It does not seem unlikely that for this *S* the wording of these three values is suggestive of the ideas of impulse-release and freedom which, after the posthypnotic suggestion, he is able to accept openly and with expressed longing.

Among the positive values seen by this *S* as associated with Los Angeles while he was still ambivalent toward that city (i.e., during the testing before the experimental hypnotic session) were the following: "having interesting work to do," "having a steady income " "people being well educated." After the posthypnotic suggestion, these earlier "pro-Los Angeles" beliefs drop out of his cognitive structure becoming unrelated to Los Angeles. With their separate stresses on work, regularity, and education, the items are representative of his usual control and suppression of instinctual energies; they represent impediments to free release, just as in his real life, "work and education" have been the major sources of control and substitutive gratification. By detaching them from Los Angeles, the *S* seems to give further evidence that the posthypnotic suggestion has freed him, at least temporarily, to acknowledge and express his usually conflicted sexuality by way of his response to the symbolically significant object of Los Angeles.

In general, the cognitive reorganization shown by this *S* offers significant support for, and adds to the plausibility of, our interpretation of the dynamic meaning of his compliance to the posthypnotic suggestion. That interpretation, in summary, is that the *S*'s chief current problem of sexual conflict has completely invested his response to the posthypnotic attitude suggestion of positive affect toward Los Angeles. Thus his comfortable and apparently guilt-free acceptance of the notion of living in Los Angeles is interpreted as due to the fact that such an advocacy of Los Angeles is, for him, an effectively disguised expression of the drives that, heretofore, have been precariously bound in an unworkable reaction-formation. That this interpretation is so is suggested not only by the test data that have been reviewed but, also, by his persisting amnesia for the posthypnotic suggestion and perhaps even by his later attainment of a somewhat better heterosexual adjustment.

SUMMARY AND CONCLUSIONS

Within the context of a general, psychoanalytically oriented theory of hypnosis there were presented two hypotheses on the nature of compliance with posthypnotic suggestions. According to the first, such compliance is viewed as facilitated by the subject's being able to interpret the content of the posthypnotic suggestion in a manner consistent with the mechanisms and affective reactions that, for him, characterize and maintain the hypnotic relationship. In the second, compliance with a posthypnotic suggestion is viewed as facilitated if that suggestion permits the subject safely to express and indulge a previously warded-off and conflicted drive. Case-record data drawn from a recent experimental study were presented which tend to confirm these two hypotheses.

In the authors' opinion, a useful area for further research would be to pursue the possibility that the present hypotheses are applicable beyond the problem of posthypnotic suggestion. Apart from hypnosis there exist many other forms of influence that seem to depend upon relationships featuring some degree and quality of identification. It is suggested that the present hypotheses might serve as a basis for investigating the processes mediating influence effects in such situations as psychotherapy and the parent-child, teacher-student, and marital relationships.

REFERENCES

1. Brenman, M., Gill, M. M., & Hacker, F. J. Alternations in the state of the ego in hypnosis. *Bull. Menninger Clinic*, 1947, **11**, 60–66.
2. Ehrenreich, G. The influence of unconscious factors on hypnotizability: A case report. *Bull. Menninger Clinic*, 1951, **15**, 45–57.
3. Freud, S. *Group psychology and the analysis of the ego.* London, Vienna: International Psycho-analytic Press, 1922.
4. Kubie, L. S., & Margolin, S. The process of hypnotism and the nature of the hypnotic state. *Amer. J. Psychiat.*, 1944, **100**, 611–622.
5. Rosenberg, M. J. Cognitive structure and attitudinal affect. *J. abnorm. soc. Psychol.*, 1956, **53**, 367–372.
6. Schilder, P., & Kauders, O. A textbook of hypnosis. In P. Schilder, *The nature of hypnosis.* New York: International Universities Press, 1956.
7. Schneck, J. Hypnosis-death and hypnosis-rebirth concepts in relation to hypnosis theory. *J. clin. exp. Hypnosis*, 1955, **3**, 40–43.
8. Schneck, J. Unconscious relationship between hypnosis and death. *Psychoanalyt. Rev.*, 1951, **38**, 271–275.
9. Weitzenhoffer, A. M. *Hypnotism: An objective study in suggestibility.* New York: Wiley, 1953.

Received November 25, 1957.

CRITIQUE AND NOTES

A TEST OF "STRAIN TOWARD SYMMETRY" THEORIES

HARRY A. BURDICK AND ALAN J. BURNES[1]

Dartmouth College

In a two-person interaction system, in which the persons are oriented toward one another and toward an inanimate object, Heider (4) has hypothesized a "strain toward balance," in which "balance" is characterized by agreement with the other person when he is positively evaluated or disagreement when he is negatively evaluated. Newcomb (7) has extended this formulation, placing the "strain toward balance" outside, as well as within, the perceptual field of the interacting organism. Using the theory of linear graphs, Cartwright and Harary (3) have in effect formalized both of these theoretical positions. While current experimental data appear to be consonant with both versions of "strain toward balance," these views have not been tested directly.

We report two studies, one done at Dartmouth College and one at the University of Michigan. In the Dartmouth study (Study I), we have measured the existence of a tension state of the organism confronted by an "unbalanced" situation. The Michigan study attempts a direct test of "strain toward symmetry" hypotheses.

STUDY I. TENSION DIFFERENCES IN "STRAINED" AND "BALANCED" SITUATIONS

The hypothesis of a "strain toward symmetry" implicitly assumes that a tension state arises within the interacting person from his perception of lack of "balance." Jordan (5), working with imaginative situations, offers some evidence that situations characterizing states of imbalance are perceived as being less pleasant than are states of balance. The following experiment employs a measure of skin resistance as an index of this underlying tension state, in line with a number of recent studies (e.g., 1, 6) which support the use of GSR as a measure of emotionality. In the present study, we attempted to create balanced and imbalanced systems for our subjects while obtaining concurrent measurement of skin resistance. The specific hypothesis was that conditions of agreement result in fewer and smaller GSR deflections than do conditions of disagreement. On the assumption that the experimenter is a positively valued "other," agreement with the experimenter is taken as a balanced state, disagreement as imbalanced.

Method

Twenty-four *S*s were volunteers from sections of the elementary psychology class at Dartmouth between December, 1956 and April, 1957. *S*s were freshman or sophomore males between 17 and 20 years old, who had little or no familiarity with GSR equipment; several tests with *S*s previously acquainted with GSR research gave no evidence that familiarity with it affected the results.

Each *S* was met by the first experimeter (E_1) and after being seated was strapped into the galvanometer leads and assured that there were no unpleasant effects from the device. *E* then withdrew into the adjoining room that contained the recording equipment, and left the *S* alone for several minutes, allowing him to believe that the equipment was operating. *S* was then introduced to the second experimenter (E_2), who took five minutes to get acquainted before presenting the test topics.

Two topics were "selected" by E_2 from a list after "joint deliberation" between E_2 and *S*: "Life after Death" and "The Draft." Each topic was discussed for approximately 15 minutes; the second immediately followed the first. Half of the *S*s discussed the draft first, half life after death. E_2 agreed with half the *S*s concerning the issue and disagreed with the remaining half. If he agreed with *S* on the first topic, he disagreed on the second. Problem order and agreement are thus varied in a factorial design, with six *S*s in each cell. *S* was allowed to choose any means of attacking the topical matter, as was E_2 within the limits of the design. E_2 often functioned as an "authority figure" by virtue of his membership on the faculty.

Measurements for GSR amplitude were obtained with an alternating current amplifier connected to a constant current isolation transformer to filter out disturbances in the electrical line. This alternating current was of uniform frequency and would show a wide range of resistance—typically from 10,000 to several million ohms. Sensitivity was maintained at constant level. The galvanic recorder furnished a record of deflections. E_1 listened to the conversation through earphones and marked each deflection with a phrase. He obtained a base line or resting resistance in the "getting acquainted" period before the topics were presented and returned all deflections (at the beginning of their decline) to this value, employing the fine potential-divider.

Data were expressed in terms of over-all deflection from the resting resistance, which was readjusted for extended shifts in resistance but not for momentary changes attributable to GSR. The large squares on the recording paper, measuring one inch by one inch, were taken as the basic units of measure. To permit comparison of agreement and disagreement phases regardless of duration, deflection scores were computed according to the following formula: sum of the amplitudes of deflections divided by the length of agreement (disagreement) phase plus time needed to reach maximum amplitude for each deflection. For each cell of the design, the basic data consisted of the difference for each *S* between the computed mean deflections under agreement conditions and the mean deflection under disagreement conditions.

---

[1] Now with the U. S. Marine Corps.

At the end of the sessions, all *S*s were asked to fill out a questionnaire concerning how they felt about the experimental situation and about E_2.

Results

For the situation to be relevant to the purpose of the study, it was essential that E_2 be evaluated positively by the *S*s. One of the questions in the postexperimental questionnaire asked accordingly: "How much do you think you'd like him (E_2) if you got to know him?" Since 22 of the 24 *S*s either said they thought they would like E_2 "very much" or "fairly well," apparently E_2 was positively valent for them. Further, 23 of the 24 volunteered to participate again. Agreement with E_2 can therefore be treated as a balanced situation, and disagreement as imbalanced.

Our major hypothesis states that there will be significantly greater deflections under disagreement conditions. Using a Wilcoxon matched-pair signed-ranks test (9), we observe a T value of 85 ($p < .05$, one-tail), in support of the hypothesis. A Kruskal-Wallis (9) analysis of variance test, however, did not produce an H value sufficiently large to reject the null hypothesis of no difference between our four cells. A Cochran test for homogeneity of variance indicates that a parametric test is inapplicable. While there is no significant difference between the variance of scores on the issues, an F test shows that the variance of scores obtained when agreement precedes disagreement is significantly greater than the variance when disagreement precedes agreement ($F = 5.61$, $p < .01$). Thus, we are unable to make any statements concerning less or greater tension resulting from our variations of problem or order.

These results are consistent with the view that disagreement with a positively valent "other" is a noxious stimulus, as would be predicted from the formulation of "a strain symmetry." It is felt that during the socialization period the individual learns that agreement with positively valent "others" is instrumental to the obtaining of positively reinforcing stimuli.

Study II. A Direct Test of "Strain Toward Symmetry"[2]

In the second study, "strain" was first induced in *S*s by disagreement with a positively valent "other," and ensuing changes in *S*s' opinions examined with respect to whether they occurred in a direction that would reduce the "strain." In a second experimental period that followed immediately, strain was induced by arousing dislike of the person who had previously been positively valent. The opinion change away from the position of the "other" is seen as an attempt to restore "balance."

Method

Subjects. Seventeen white male sophomore or junior transfer students at the University of Michigan served as *S*s. The men had been living together in a house rented by the research project for approximately five weeks holding two-hour meetings one evening a week. Two additional hours per week were spent filling out questionnaires and being interviewed. Prior to their arrival, they had all been strangers to one another. They included both veterans and nonveterans and members of the three major religious groups. They were ignorant of the purpose of the present research, which comprised but a small part of the total project.

Procedure. Prior to the arrival of the "speaker" (E_1), *S*s were required to fill out the first set of questionnaires. Having read a case study of a juvenile delinquent (2), *S*s indicated their opinions concerning whether the boy personally, or the boy's environment, should be considered responsible for a crime which he committed. They also indicated their opinions concerning the proper dress for college men on campus. Both opinions were obtained by having *S*s record their first and second choices of positions.

E_1 then was introduced as a member of the University's Psychology Department who would speak to them concerning the two topics.[3] E_1 spoke pleasantly for an hour about the case which *S* had read, arguing that the environment should be considered totally responsible for the boy's crime. At the end of the hour, *S*s again filled out the two opinion questionnaires and a five-alternative scale about how well they liked the speaker.

E_1 returned to the group when the *S*s had finished filling out the questionnaires and announced that, contrary to his original plans, he had no intention of speaking about college clothing. Instead, he spent the remaining hour berating the *S*s for their docility and submissiveness when confronted by an authority. He discussed neither juvenile delinquency nor college clothing, but attempted to make himself disliked with no reference to the topics. At the end of this hour, *S*s again filled out the questionnaire on the issues and the one on liking for the speaker. When the deception was explained to *S*s at the close of the following academic year, there were indications of some suspicion by then on the part of some *S*s.

The questions were repeated during the 12th and 15th weeks of the semester. During the eighth week, *S*s filled out the need for affiliation measure described by Shipley and Veroff (8). In this test, S writes imaginative stories about some standard pictures. The person who scored these stories for evidence of the "need state" was not familiar with the experimental design used here.

---

[2] This study was part of a broader investigation sponsored by the Rockefeller Foundation. It was conducted at the University of Michigan during the years 1954–1956, under the direction of T. M. Newcomb.

[3] Elton McNeil of the Psychology Department of the University of Michigan performed the role of E_1.

Results

Success of experimental manipulations. To meet the conditions for testing our hypotheses, it was necessary that Ss be attracted to E_1 during the first hour of our experiment. The questionnaire results indicate that they were. At the end of the first hour, five Ss said they liked E_1 "very much," nine said they "liked him," two said "neither liked nor disliked him," and the one remaining person said that he "disliked him." It was further essential that the Ss change their opinions of E_1 by the end of the second hour. Comparison of the questionnaires given at the close of each period indicate that 14 of the 17 Ss liked E_1 less at the close of Phase 2 than they had at the close of Phase 1. (The probability of such a change occurring by chance is less than .01.)

Opinion change and "strain toward symmetry." At the end of the first hour, eight Ss indicated that they had changed their opinion regarding the issue, all of them taking a position closer to E_1's position than they had originally. A sign test indicates that the probability of such change is .004. But at the end of the second hour, six Ss had changed their opinion from the position they had taken at the end of the first hour, in spite of the fact that E_1 had said nothing further on the topic. Five of the six changed in a direction away from the position taken by E_1. A sign test indicates that the probability of such a change is .109. On the "dress on campus" issue, which was not discussed, no significant changes of opinion were observed.

The repeated measures taken in subsequent weeks do not indicate any significant or orderly change of opinion. During the last week of experimentation, Ss indicated that they still were unfavorably predisposed to E_1, though some shift toward the "neither liked nor disliked" alternative was observable.

The results in general appear to support our hypotheses, although it is obvious that a large number of the Ss did not change their opinions. The question arises in what respects these persons might differ from the "changers." Data on the S's need for affiliation were examined from this perspective.

Need affiliation and opinion change. Ten Ss wrote stories with no evidence of need for affiliation; seven wrote stories with affiliation imagery. The Ss were divided into two groups, with ten persons relatively low and seven relatively high of the measure. The prediction was that those high in need for affiliation would show most change in response to the situation in which expected change depended upon a "liking." Table 1 presents the comparisons (Fisher exact test) for the two motivational groups at each phase of the experiment. The three persons who were originally at the extreme environmental end of the judgmental continuum prior to E_1's appearance are excluded from the first comparison on the grounds that they could not possibly move toward this position. They could, however, move away from him and so are included in the second comparison. Table 2 presents this analysis. The results of the comparisons indicate that the affiliation motive is related to change in opinion under the circumstances of the present study.

TABLE 1
THE RELATIONSHIP OF NEED FOR AFFILIATION TO OPINION CHANGE IN PHASE 1[a]

Need for Affiliation	Direction of Change: Toward E_1	Direction of Change: Not Toward E_1
High	6	0
Low	2	6
Exact test of significance	$p < .025$[b]	

[a] Three Ss initially in full agreement with E_1 are excluded.
[b] One-tailed test of significance.

TABLE 2
THE RELATIONSHIP OF NEED FOR AFFILIATION TO OPINION CHANGE IN PHASE 2

Need for Affiliation	Direction of Change: Away from E_1	Direction of Change: Not Away from E_1
High	5	2
Low	1	9
Exact test of significance	$p < .025$*	

* One-tailed test of significance.

SUMMARY

Two studies have been reported which offer some support to the "strain toward symmetry" hypothesis. GSR deflections were found to be associated with disagreement with a positively valent experimenter. Two topics were used as matters for disagreement, but we were unable to discover significant differences in the deflections associated with the topics used.

In the second experiment, the attractiveness of the person serving as a source of communication was directly manipulated. Subjects who reported that they "liked" him tended to change their opinions toward more agreement with him. When they were subsequently induced to "dislike" him, they tended to change their opinions toward greater disagreement. Evidence is presented that this change of opinion is related to need for affiliation.

REFERENCES

1. Bitterman, M. E., & Holtzman, W. H. Conditioning of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615–623.
2. Burdick, H. A. The relationship of attraction, need achievement and certainty to conformity under conditions of a simulated group atmosphere. Unpublished doctoral thesis, Univer. of Michigan, 1956.
3. Cartwright, D., & Harary, F. Structural balance: A generalization of Heider's theory. *Psychol. Rev.*, 1956, **63**, 277–293.
4. Heider, F. Attitudes and cognitive organization. *J. Psychol.*, 1946, **21**, 107–212.
5. Jordan, N. Behavioral forces that are a function of attitudes and of cognitive organization. *Hum. Relat.*, 1953, **6**, 273–287.
6. McGinnies, E. M. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
7. Newcomb, T. M. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, **60**, 393–404.
8. Shipley, T. E., Jr., & Veroff, J. A. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, **43**, 349–356.
9. Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Received June 28, 1957.

MARITAL SATISFACTION AND HUSBANDS' AND WIVES' PERCEPTION OF SIMILARITY IN THEIR PREFERRED FREQUENCY OF COITUS[1]

PAUL WALLIN AND ALEXANDER CLARK

Stanford University

The research reported here supports the hypothesis that there is a positive association between an individual's favorable or unfavorable valuation of another and his perception of the other as like or unlike himself. A number of studies (1, 3, 4, 5, 6, 7) in recent years have been concerned with this question centrally or peripherally. Two of the studies (4, 7) have dealt with the question in the context of marriage, investigating marital satisfaction as a condition of husbands' and wives' perception of their spouses as similar to themselves in selected personality attributes.[2] The present study is also concerned with marital satisfaction as a condition of perception of similarity but differs from previous research in that it considers perceived similarity with respect to the preferred frequency of intercourse of marital partners.

There is in our culture a widespread belief—which may or may not be warranted—that similarity of persons or groups is necessary for positive feeling and harmony in interpersonal or intergroup relationships. It is thought, for example, that common interests and similarity of background are vital for a satisfactory marriage. Hence, it might be expected that persons would tend to perceive greater similarity between themselves and those whom they liked and with whom they had amicable relations than would be the case with persons whom they disliked and with whom they were in conflict. Perception is here conditioned by the culturally given explanation for liking or disliking, harmony or conflict.

A second basis for the expectation that greater similarity would be imputed to liked than to disliked persons is that real differences are probably more often muted in interaction with the former and accented in interaction with the latter. In liking someone we are motivated to perpetuate the relationship. The fewer the actual differences perceived in interests, values, personality attributes—particularly those dissimilarities which it is believed might threaten the relationship—the more secure we can feel about it. Moreover, the lesser perception of differences in such a relationship is facilitated by compromise of differences which narrows the extent of apparent dissimilarity. On the other hand, in disliking someone we are less disposed to compromise differences. Differences may even be exaggerated to rationalize the dislike or the anticipated termination of the relationship.

---

[1] This is one of a group of articles reporting the findings of a study dealing with questions relating to the sexual behavior of men and women before and after marriage. Acknowledgment is due for support given in the study by the Rockefeller Foundation and the Social Science Research Council. The research was also supported in part by a grant from the Committee for Research in Problems of Sex, National Academy of Sciences—National Research Council. The senior author is indebted to Ernest W. Burgess with whom he collaborated on the long-term investigation of engagement and marriage in the course of which the data of the sex study were collected. For their work on various phases of the study thanks are due to Fred Chino, Jan Howard, and Don Mills.

[2] It is assumed that the more harmonious and satisfactory a marriage the more positive the feeling which men and women have for their marital partners.

In terms of the content of the present research, then, it might be expected that maritally satisfied persons, more often than the dissatisfied, would tend to perceive their spouses' preferred frequency of coitus as similar to their own. It may, of course, be legitimately objected that persons who are satisfied with their marriage more often perceive this similarity because they are, in fact, more similar in this regard than dissatisfied couples. This objection is met in the analysis that follows.

SAMPLE AND DATA

The persons whose experiences are the basis for this report participated in an extensive research dealing with many questions relating to engagement and marriage. Since this study is treated in detail elsewhere (2) it need only be described briefly here.

In its first stage it was concerned with 1000 engaged couples. Approximately three-fifths of the couples were restudied after a few years of marriage. At that time, husband and wife independently of one another and under supervision, filled out questionnaires that yielded the information used for the study reported in this article. This phase of the study was completed in 1946.

The couples were volunteers and received no payment. They were white, native born, and predominantly residents of metropolitan Chicago. Only a fifth of the men and about a third of the women lacked some college education. About a fourth of the men and a tenth of the women were at the graduate or professional level of education. Somewhat more than half the couples were Protestant, the others being Jewish or Catholic or having no religious affiliation. The average age at the time of marriage was 25.7 for the men and 23.9 for the women.

About three-fourths of the couples had been married between three and five years. A small percentage had been married under three years, and the marriages of the remainder were of six to nine years' duration. Consequently, the findings presented below are for the early years of marriage. A follow-up study is now being made of these couples in the middle years of marriage.

The data used for determining *perceived and actual similarity* of husbands and wives in their preferred frequency of coitus were their responses to the following two questions:

(*a*) About how many times *per month* would you prefer to have sexual intercourse?

(*b*) About how many times *per month* do you think your wife (husband) would prefer to have sexual intercourse?

To allow for a margin of chance error in estimates of a monthly rate, husbands and wives were treated as perceiving their marital partners as similar if their own preferred frequency differed by less than three from the preferred frequency they imputed to their marital partners.

The criterion of actual similarity of couple members was whether they were similar or different as judged by their own reports of their preferred frequency of coitus.[3] Here again, similarity was defined as a difference of less than three in the indicated monthly rate.

Finally, the *marital satisfaction* of husbands and wives was measured by their independent responses to a series of items indicative of general satisfaction or over-all contentment with the marriage and the marriage partner. (For example: How satisfied, on the whole, are you with your marriage?) The items (and their weights), responses to which determined the marriage satisfaction scores of the men and women, are reported elsewhere (2, pp. 488–489 and 809–810).[4] In the analysis presented below, the distributions of the marital satisfaction scores of husbands and wives were dichotomized with the breaks made as close to the medians as the distributions permitted. Scores above these cutting points are designated as "high" (indicative of greater marital satisfaction), scores below the cutting points are designated as "low" (indicative of lesser marital satisfaction).

FINDINGS

The findings for the initial test of the hypothesis are positive. As can be seen in Table 1, a significantly larger percentage of husbands and wives who score high on marital satisfaction than of those who score low report their spouses' preferred monthly frequency of coitus as similar to their own.[5]

---

[3] The accuracy of reports of persons' own preference is discussed in a paper being readied for publication which deals with factors associated with women's preferred frequency of marital coitus.

[4] Items 22, 24, and 26 were not used for the marriage satisfaction scores of the present study. The scoring of the items was somewhat different from that previously reported (2, pp. 809–810) but yields essentially the same ranking of persons on marital satisfaction. In the case of the wives, the marital satisfaction scores correlated .78 with scores on another series of items relating to marital happiness. The parallel correlation for the husbands was .73.

[5] The total number of couples reported on in Tables 1 and 2 is less than that of the original sample (604) because couples were excluded if either husband or wife failed to report their own preferred frequency or that of the marital partner. In some instances, couples were

TABLE 1

MARITAL SATISFACTION SCORES AND PERCEPTION OF SPOUSES' COITAL PREFERENCE

Marital Satisfaction Score	Husbands' Perception of Wives' Preference: Similar[a]	Dissimilar[b]	Total No.	Wives' Perception of Husbands' Preference: Similar[a]	Dissimilar[b]	Total No.
	%	%		%	%	
High	76.8	23.2	250	73.2	26.8	205
Low	52.2	47.8	209	58.3	41.7	254
	$\chi^2 = 30.73$ $P < .001$			$\chi^2 = 11.06$ $P < .001$		

[a] Differing by less than two from own preference.

[b] Differing by more than two from own preference.

TABLE 2

MARITAL SATISFACTION SCORES, PERCEPTION OF SPOUSES' COITAL PREFERENCE, AND ACTUAL SIMILARITY OF PREFERENCE[a]

Marital Satisfaction Score	Class I: Husbands and Wives Have Similar Preference[b]	Class II: Husband's Preference Higher[c]	Class III: Wife's Preference Higher[c]
Husbands[d]			
High	84.9% ($N = 139$)	55.4% ($N = 74$)	89.2% ($N = 37$)
Low	67.4% (95)	28.0 (75)	61.5 (39)
Wives[e]			
High	76.1% (113)	59.3% (54)	84.2% (38)
Low	66.9 (121)	43.2 (95)	68.4 (38)

[a] Percentages in the table indicate the proportion of persons in each cell who perceive their spouses' preferred frequency as similar to their own, i.e., as differing by less than three times per month.

[b] Couples in which by self-report husband and wife differ by less than three times per month.

[c] Couples in which by self-report husband and wife differ by three or more times per month.

[d] Chi squares for Class I, II, and III husbands are respectively 10.05, 11.49, 7.89. For each of these, P is .001 or higher.

[e] Chi squares for Class I, II, and III wives are respectively 2.44, 3.52, 2.62. Using the one-tail test, only 3.52 achieves the .05 level of significance.

To rule out the possibility that maritally satisfied husbands and wives in fact are more congruent in their preferred frequency of coitus, the association between marital satisfaction and perception of similarity was examined with couples classified by the *actual* preference of husband and wife. As pointed out earlier, the measure of the latter was husbands' and wives' own reports of their preferred frequency. With this criterion, couples were divided into three groups as shown in Table 2. Class I couples are those whose members differed by less than three in their self-reports of their monthly preference.[6] Couples whose members differed by three or more in their reports of their own preferred frequency were placed in Class II or III depending on whether it was the husband or wife who had the higher preferred frequency.[7] These two classes have been kept separate because of the finding in an earlier study[8] that in couples in which the wife's preference exceeds that of the husband (Class III) each couple member is influenced in reporting the other's preference relative to his (her) own by the cultural norm that marital partners should have the same sex drive and that when they differ the husband's drive should be the stronger. The influence of this norm on Class III husbands and wives conceivably could have counteracted the hypothesized effect of differences in marital satisfaction on the perception of similarity or dissimilarity of marital partners. For this reason the data for Class II and III couples were analysed separately with the assumption that the findings for Class II persons would be more decisive for the hypothesis being tested.

Table 2 reveals that in all three classes of couples a greater proportion of husbands and wives who are high on marital satisfaction than of those who are low report their spouses' preferred frequency of coitus as similar to their own. The differences for the husbands are significant at the .01 level or better. In the case of the wives the differences are all in the predicted direction, but only one of them is statistically significant.

The results of the analysis presented in Table 2 are firmly consistent for the husbands with the hypothesis that marital satisfaction is positively associated with perception of similarity to marital partner in preferred frequency of coitus. This association holds for marriages in which couple members are presumed to be similar in preference and for those in which they differ. The results for the wives are consistent with the hypothesis but less reliably so.

SUMMARY

The study reported here tested the hypothesis that maritally satisfied persons tend to perceive

---

excluded because husband or wife responded qualitatively (e.g., "often") or by specifying a broad range of frequency (e.g., 10 to 20).

[6] The findings subsequently reported were essentially the same when Class I couple members were grouped in terms of no difference at all in self-reports of husbands and wives and in terms of a difference of either 1 or 2.

[7] The findings subsequently reported were essentially the same when Class II and III couple members were grouped as differing by 3, 4, 5, 6, 7, 8, or more in self-reports. In Class II, the magnitude of dissimilarity between husbands and wives in self-reports was greater for the maritally dissatisfied persons than for the maritally satisfied.

[8] Unpublished manuscript by P. Wallin and A. Clark, "Cultural norms and husbands' and wives' reports of their marital partners' preferred frequency of coitus relative to their own."

their marital partners as similar to themselves in preferred frequency of coitus more often than do the maritally dissatisfied. The data for the study were obtained by questionnaires independently answered by husbands and wives. The measure of perceived similarity was the congruence between persons' reports of their own and their marital partners' preferred monthly frequency. The preferences of couple members were presumed to be similar in fact if their own reports of their preferred frequency represented them as similar. The findings for the husbands were reliably consistent with the hypothesis but less certainly so for the wives.

REFERENCES

1. Bieri, J. Changes in interpersonal perceptions following social interaction. *J. abnorm. soc. Psychol.*, 1953, **48**, 61–66.
2. Burgess, E. W., & Wallin, P. *Engagement and marriage.* Philadelphia: Lippincott, 1953.
3. Davitz, J. R. Social perception and sociometric choice of children. *J. abnorm. soc. Psychol.*, 1955, **50**, 173–176.
4. Farber, B. An index of marital integration. *Sociometry*, 1957, **20**, 117–134.
5. Fiedler, F. E., Warrington, W. G., & Blaisdell, F. J. Unconscious attitudes as correlates of sociometric choice in a social group. *J. abnorm. soc. Psychol.*, 1952, **47**, 790–796.
6. Fiedler, F. E. Assumed similarity measures as predictors of team effectiveness. *J. abnorm. soc. Psychol.*, 1954, **49**, 381–387.
7. Preston, M. G., Peltz, W. L., Mudd, Emily H., & Froscher, H. B. Impressions of personality as a function of marital conflict. *J. abnorm. soc. Psychol.*, 1952, **47**, 326–336.

Received August 7, 1957.

A "PURE" MEASURE OF PERCEPTUAL DEFENSE UNCONTAMINATED BY RESPONSE SUPPRESSION

ANNE MATHEWS[1] AND MICHAEL WERTHEIMER[2]

University of Colorado

Many methodological difficulties associated with perceptual defense studies have been recognized and corrected, but it has been particularly difficult to isolate and control the variable of response suppression. As Eriksen (3, 4, 5), Postman (8), and others (e.g., 6, 9) have pointed out, until a study demonstrates that perceptual defense cannot be accounted for by response suppression, the existence of true perceptual defense must be doubted. The aim of this paper is to present such a study.

METHOD

An experimental design similar to Blum's (1) was devised to permit measurement of perceptual defense in a situation of tachistoscopic recognition, with a separate measure of the amount of response suppression also available.

In individual sessions, *S*s were given the task of identifying words flashed tachistoscopically on a screen. The particular tachistoscopic setting used was well below 50% recognition threshold, but it had previously been demonstrated (on a group of "calibrating" *S*s) to result in slightly above chance recognition. Each *S* was handed a list of eight words and was told by *E* that every word flashed on the screen would be some word from that list. Four of the words on the list were emotional and had been chosen on the basis of that *S*'s reactions to a word-association test given in an earlier individual session (details of this procedure, and of other aspects of the study, can be found in Mathews [7]). The other four words were neutral[3] and were paired with the emotional words for length, approximate structure, and relative frequency of occurrence in the Thorndike-Lorge General Word Count. During the tachistoscopic task, therefore, *S* had a total of eight possible alternative hypotheses about which word would be flashed, but he had no prior knowledge for predicting *which* word would be presented on any one particular flash. Each *S* was confronted with 32 flashes or presentations of a word.

Unknown to *S*, however, only *two* emotional

---

[1] Now with Educational Testing Service, Princeton, New Jersey.

[2] The completion of this paper was made possible by Wertheimer's participation in the Behavioral Science Conference at the University of New Mexico, Summer, 1957, under Project AF49(638)-33 sponsored by the Behavioral Science Division, Air Force Office of Scientific Research, ARDC, Paul Walter, Jr., principal investigator.

[3] The criterion for "neutrality" here was simply the mutual agreement of the authors that the word was probably lacking in emotional connotations for most people. Words such as Sit, Chair, North, Four, Day, and Many were used. As will be seen below, positive results were obtained in spite of the weakness of this technique (compared, e.g., with a technique in which a word association method is used with each individual *S* for the assignment of neutral words, but which creates the problem of how to match emotional and neutral words structurally).

and *two* paired neutral words were *actually* flashed; each of these four words was presented eight times in random order. The other four words were *never* flashed. Thus, this procedure permitted the abstraction of four categories for classifying *S*'s responses (two words being represented in each category):

Present (Words actually presented)	Absent (Words not presented)
a Emotional words	*c* Emotional words
b Neutral words paired with words in *a*	*d* Neutral words paired with words in *c*

Since people with hysteric tendencies might be considered likely to be repressors (2) and hence to show perceptual defense, this technique was used with 20 *S*s scoring high on the hysteria (*Hy*) scale of the MMPI—selected from a group of 271 students (7). A group of 20 *S*s scoring high on the psychasthenia (*Pt*) scale of the MMPI was also chosen for the experiment because it was considered that they might be hypersensitive to emotional material and, hence, might show perceptual vigilance (2).

The primary aim of this study was to obtain a single score for each individual which would be an index of "pure" perceptual defense—an index uncontaminated by any tendencies for response suppression. By correcting indices obtained from *a* and *b* by indices obtained from *c* and *d*, this aim could be accomplished. These indices were obtained as follows:

Through the two "Stimulus Absent" categories, *c* and *d*, an index of "pure" response tendencies is made available, since no stimulus corresponding to these four words was ever presented. One would expect that if *S* had no differential response tendencies, then 50% of his calls for words not actually presented would be calls of emotional words and 50% would be calls of neutral words.

This interpretation involves the crucial assumption of equal probability of calls for emotional and neutral words in the "Stimulus Absent" category. It might be argued that when *S* calls out a word that has *not actually been presented*, the nature of his error cannot be considered entirely independent of the stimulus which actually *was* presented. Thus the frequency of these two possible types of "misperceptions" (calls for emotional absent and neutral absent words) might be influenced by the relative similarity of the present stimuli to the absent stimuli. The argument is undoubtedly sound theoretically. However, a detailed analysis of calls actually made yielded the result that the variable of similarity between Present and Absent words had virtually no effect on our *S*s' "misperceptions." Aside from this analysis, there are further grounds for the assumption of an equal distribution of calls between Categories *c* and *d* if there is no response suppression. A priori, there is no reason to expect greater similarity of a presented word to an absent neutral than an absent emotional word; initial pairing of words in Category *c* with those in Category *d* had been made on the basis of similarity, so that a presented word would be about equally similar to the absent emotional and neutral words. Thus, in most instances at least, if a word actually presented to a given *S* (e.g., *suck*) resembled a word in Category *c* (e.g., *self*), it also resembled the paired word in Category *d* (e.g., *sale*). It can therefore be argued that any degree of undercalling (or overcalling) of emotional words as compared to neutral words can be interpreted as the degree of response suppression (or response sensitivity) manifested by *S*.

The expected number of calls of emotional words in Category *c* can therefore be calculated for each *S* by the formula $M = Np$, where N is the number of absent words called, and p is, by assumption, .5. Likewise, the standard deviation of total calls in the two "Stimulus Absent" categories can be calculated by the formula $\sigma = \sqrt{Npq}$.

In order to express each *S*'s deviation from a theoretically expected frequency of calls for emotional words in Category *c*, a "*z*" score $(X - M/\sigma)$ was used. The "*z*" score indicates both the extent and the nature of *S*'s response tendencies—a plus "*z*" score indicating response sensitivity, a zero "*z*" score indicating no differential tendencies, and a minus "*z*" score indicating response suppression.

However, in Categories *a* and *b* (words actually presented), relative calling of the two categories provides an index of *both* response tendencies *and* perceptual tendencies. The assumption that total "Stimulus Present" calls will be equally distributed between emotional and neutral categories (if neither response suppression nor perceptual defense occurs) still holds, since these words were in fact presented an equal number of times. Following the same procedure as described above, then, a "*z*" score can also be calculated for each *S* which represents his deviation from a theoretically expected frequency of calls for emotional words actually presented. Such a "*z*" score is an index that confounds both perceptual and response tendencies.

In order to obtain an index of "pure" perceptual tendencies, each *S*'s "*z*" score for "Stimulus Absent" words was subtracted from his "*z*" score for "Stimulus Present" words. Interpretation of the resulting difference "*z*" score for each individual can now be made on the basis of perceptual processes alone—since this is an index from which

TABLE 1

INDIVIDUAL "z" SCORES FOR *Hy* AND *Pt* SUBJECTS (All scores are positive unless otherwise indicated)

Hy Group (Difference "z" scores)	*Pt* Group (Difference "z" scores)
1.92	−.48
−1.00	.99
−.50	−1.45
−.77	−.46
−.51	−1.88
.42	3.51
−.27	.80
−1.34	0.00
−.85	−.07
0.00	−.91
−.38	−1.84
.20	0.00
0.00	−.83
0.00	.80
−2.24	.09
0.00	0.00
0.00	−.27
−2.33	0.00
.25	−1.55
−.67	.28
$M = -.404$	$M = -.164$
CR = 1.94 ($p = .026$)	CR = .61 ($p = .271$)

the response tendencies of *S* have presumably been partialled out. Two assumptions are required to make this interpretation entirely tenable: namely, that response suppression (or response sensitivity) occurred to the same extent with the words actually presented as with those that were not presented,[4] and that response suppression effects and perceptual defense effects summate.[5]

In brief, the statistical procedure that was used in the study can be summarized as follows:

Category *a*		Category *c*		
Present Emotional "z" score (Perceptual + Response Tendencies)	−	Absent Emotional "z" score (Response Tendencies)	=	Difference Score (Pure Perceptual Tendencies)

RESULTS

Tables 1 and 2 present the results in the form of distributions of the difference "z" scores for both *Hy* and *Pt* *S*s (Table 1) and in the form of group means and standard deviations (Table 2). The difference "z" scores are the result of algebraically subtracting the response suppression index from the contaminated index. As can be seen from the Tables, the *Hy* *S*s did manifest a perceptual defense effect over and above their response suppression effect. This effect was significant at better than the .03 level of significance.

TABLE 2

GROUP "z" SCORE MEANS AND SIGMAS

Group	Present Emotional "z"—from Categories *a* and *b*		Absent Emotional "z"—from Categories *c* and *d*		Difference "z"	
	M	σ	M	σ	M	σ
Hy ($N = 20$)	−.459	.96	−.056	.32	−.404	.91
Pt ($N = 20$)	−.386	.89	−.223	.73	−.164	1.17

As can also be seen from Tables 1 and 2, the *Pt* *S*s did not manifest the predicted perceptual vigilance (which would have been represented by a mean "z" score in the plus direction). Nor, however, did they show a significant perceptual defense effect, although there was a statistically insignificant trend in that direction. Since a t test showed the two groups did not differ from one another significantly (Diff. = .240, $t = .70$ $p > .10$), the results from both groups were combined, with the finding that twice as many *S*s exhibited perceptual defense (a negative "z" score) as exhibited perceptual vigilance (a positive "z" score).

SUMMARY

Using a tachistoscopic recognition technique, an experiment was designed that provided measures both of "perceptual defense" and of response suppression. The "perceptual defense" measure could be corrected by the response suppression measure to yield a "pure" perceptual defense measure. With this method, it was found that *S*s scoring high on the *Hy* scale of the MMPI demonstrate a significant perceptual defense effect over and above a response suppression effect, when confronted by words preselected so as to be emotional for each *S* individually. It is therefore concluded that perceptual defense effects cannot be accounted for entirely on the basis of response suppression.

REFERENCES

1. BLUM, G. S. Perceptual defense revisited. *J. abnorm. soc. Psychol.*, 1955, **51**, 24–29.
2. ERIKSEN, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor Anxiety Scale and the hysteric and psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135–137.
3. ERIKSEN, C. W., & BROWNE, C. T. An experimental and theoretical analysis of perceptual defense. *J. abnorm. soc. Psychol.*, 1956, **52**, 224–230.
4. ERIKSEN, C. W. Personality. In P. R. Farnsworth &

---

[4] This assumption seemed warranted in our study, since a counterbalanced technique was used to assign the four "most emotional" words for a given *S* to the "Stimulus Absent" and "Stimulus Present" category.

[5] Such a summation assumption may be oversimplified, but more complex comparable assumptions of an interactive kind would still yield results similar to those obtained.

Q. McNemar (Eds.), *Annual review of psychology.* Palo Alto, Calif.: Annual Reviews, 1957. Vol. 8.
5. Garner, W. R., Hake, H. W., & Eriksen, C. W. Operationism and the concept of perception. *Psychol. Rev.*, 1956, **63**, 149–159.
6. Howes, D. H., & Solomon, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229–234.
7. Mathews, Anne. A "pure" measure of perceptual defense. Unpublished master's thesis, Univer. of Colorado, 1957.
8. Postman, L. Perception, motivation, and behavior. *J. Pers.*, 1953, **22**, 17–31.
9. Whittaker, E. M., Gilchrist, J. C., & Fischer, J. W. Perceptual defense or response suppression? *J. abnorm. soc. Psychol.*, 1952, **47**, 732–733.

Received August 20, 1957.

PERCEIVED SIMILARITY AND THE PROJECTION OF SIMILARITY: THE INFLUENCE OF VALENCE[1]

ANTHONY J. SMITH
University of Kansas

Over a period of years a system for conceptualizing interpersonal relations has been developed by Heider. It was first described in an article concerned with cognitive configurations and the conception of a balanced state (1). A more complete description of the system is now available (2).

The concept of structural balance can be stated in terms of a triadic relationship among a person "p" (who represents the subject), a second person "o," and an act, property, or characteristic "x." Though it is possible to extend the notion of structural balance to simpler or more complex situations, this paper is concerned only with the triad. Attention will be further confined to the triad in which one type of relationship exists among the variables: the relationship of liking or disliking, e.g., p L o (p likes o) or p∼Lo (p does not like o). The concept of balance may be illustrated very briefly as follows: When the subject p likes person o and o subscribes to opinion x, a force is generated such that p tends to agree with opinion x. If p agrees with x, the three elements p, o, and x, are in a state of balance. If p does not agree with statement x, an imbalanced state arises.

Horowitz, Lyons, and Perlmutter (3) have tested Heider's concept of balance in a field study of discussion groups. They found that *S*s are more acceptant of the acts of highly valued persons than they are of the acts of persons who are less valued. Further, they discovered that the *S* attributed to other accepted persons an acceptance of these same acts and were more likely to attribute to disliked persons a rejection of the acts. These results were consistent with the investigators' predictions.

Jordan (4) confronted *S*s with a series of relatively abstractly stated triadic relationships that were predicted to vary along a balance–imbalance continuum. Ratings of pleasantness were obtained from *S*s as indicators of degree of balance. In general, situations predicted to be balanced were found to be more pleasant than those predicted to be unbalanced, though some systematic exceptions were encountered.

In a later investigation, Smith (5) asked *S*s to examine partially completed test booklets (Revised Allport-Vernon) of hypothetical persons whose answers were either consistently similar or dissimilar to the *S*'s own. On the basis of the information provided, *S* completed the test booklets as he thought the hypothetical individuals would and later indicated his degree of acceptance of these individuals as work and leisure-time associates.

It was found that high valence (acceptance) was attributed to persons who were similar, whereas zero valence was attributed to those who were dissimilar or less similar. Moreover, greater additional similarity was projected on the more accepted persons than on the less accepted persons. It was reasoned on the basis of Heider's theory that this was a consequence of the differential in valences in the hypothetical o's.

In the study cited, the value scale employed to determine the extent of projection of similarity was that having the lowest valence for the *S*. The present investigation was designed in part to test the hypothesis that greater similarity is projected on similar persons than on dissimilar persons for both high and low valence materials.

A more important question concerns the differential influence of the valence of the test scale. It is hypothesized that the difference between the degree of similarity projected on a hypothetical similar individual and that projected on a hypothetical dissimilar individual is greater for

---

[1] This study was supported by a grant from the University of Kansas Research Fund. The author wishes to acknowledge the assistance of Harrison Madden and Ronald Sobol.

high valence material than for low valence material.

METHOD

Subjects

*S*s in this study were 104 students enrolled in undergraduate courses in psychology and human relations. Forty-nine *S*s were assigned to the low valence test scale group (including the 28 students reported in the original study [5]); the remaining 55 were assigned to the high valence test scale group. All *S*s were required to perform the task twice under conditions of high and low similarity respectively. Forty-nine *S*s in the high valence group were also required to respond to an intermediate degree of similarity.

Test Instrument

During the first two weeks in the semester, the Revised Allport-Vernon Scale of Values was administered as a routine part of the courses in general psychology, personality and cases in human relations from which the experimental *S*s were drawn. The test booklets were scored by research assistants and the scales on which the students received the highest and lowest scores were determined.

The procedure by means of which hypothetical similar and dissimilar persons were created has been described in detail elsewhere (5). It involved preparing partially completed experimental test booklets based on the answers of the *S*. Information concerning the scale on which a particular *S* received the highest score (high valence test scale group) or the lowest score (low valence test scale group) was systematically eliminated from the experimental test booklet provided for him.

The information that was presented in the test booklet was designed either (*a*) to duplicate the answers of the *S*, thereby creating an hypothetical similar person or (*b*) to provide the maximum systematic difference from *S*'s answers, presumably yielding an hypothetical dissimilar person or (*c*) to yield a moderate systematic resemblance to the *S*'s answers resulting in a moderately similar hypothetical person.

Instructions

Six to ten weeks following the administration of the original test, *S*s were met either individually or in small groups. They were reminded that in their everyday life they were always required to make judgments concerning characteristics of persons about which they had no direct knowledge on the basis of other information that they did possess. This experiment was designed to test their accuracy in this function. They would be provided with two or three partially completed test booklets filled out by other persons. Their task was to study each booklet individually until a clear impression of the original person emerged and then to complete the test booklet as the original person had.

ANALYSIS AND RESULTS

Similarity and the Projection of Similarity

The amount of similarity projected on each of the two or three hypothetical persons was defined as being inversely related to the absolute difference between the test score earned by the *S* on the critical scale and the score on the same scale in the experimental test booklet as completed by the *S*. This latter score was determined solely by *S*. Thus, the smaller the difference the greater the similarity attributed to the hypothetical person.

TABLE 1

MEANS OF DIFFERENCES IN AMOUNT OF SIMILARITY PROJECTED ON HYPOTHETICAL SIMILAR INDIVIDUALS AND HYPOTHETICAL DISSIMILAR INDIVIDUALS FOR TWO LEVELS OF TEST SCALE VALENCE

Valence	N	Mean Difference	σ_M	CR	P
Low Valence test scale	49	−4.78	1.28	3.73	.0002
High Valence test scale	55	−9.12	1.49	6.12	<.0001

The results of the analysis of low valence test scales are found in Table 1. The average difference between the similarity attributed to similar individuals and that attributed to dissimilar individuals is found to be −4.78 with a standard error of 1.28. In testing the significance of this difference from zero, a critical ratio of 3.73 is found with a corresponding probability of .0002.

An equivalent analysis for the high valence test scales is also reported in Table 1. The mean difference between amount of similarity attributed to similar individuals and that attributed to dissimilar individuals is −9.12, with a standard error of 1.49. The resulting critical ratio is 6.12, with a probability of <.0001.

A second form of analysis was also used in testing the first hypothesis. For 49 *S*s, scores were available for three levels of similarity on the high valence test scale. If the hypothesis were to be supported, the absolute difference between own score and projected score should be lowest for similar test booklets, intermediate for the booklets showing moderate similarity and highest for the dissimilar test booklets. The three absolute difference scores for each person were ranked and the coefficient of concordance computed for the resulting sets of ranks. A coefficient of .41 was obtained. χ_r^2 was found to be 23.64, with a probability value of less than .001.

Differential Effects of Valences of the Test Scales

The second hypothesis was tested very simply by determining whether the mean difference for the high valence test scale (−9.12) was significantly greater than the mean difference for the low valence test scale (−4.78). The standard error of the difference between these mean differences is found to be 1.96. The resulting critical ratio is 2.21 with a probability of $<.03$.

While the experiment was in progress, and during the discussions that followed the experimental sessions, questions concerning the test booklets were raised either publicly or privately by only ten *S*s. Thus it would seem that most *S*s remained naive throughout the experiment. These questions ranged from "Are these real people?" to "Was one of these mine?" On the assumption that these suspicions might have influenced the results, the data were reanalyzed excluding these ten *S*s. A critical ratio of 2.26 ($p < .05$) was obtained for the remaining 94 *S*s.

DISCUSSION

The theory of structural balance proposes that under the condition pLo, balance is attained when the remaining two elements of the triad are of equivalent valences, that is pLx, oLx, or p~Lx, o~Lx. Conversely, under either of the latter conditions, balance is achieved when pLo. This is precisely the situation that exists in the present experiment with respect to the hypothetical similar individual. When the *S* inspects the partially completed test booklet, each item establishes either the condition pLx, oLx or the condition p~Lx, o~Lx. Consequently, each item contributes an increment to the force in the direction of pLo.

When the *S* is obliged to complete the test booklet, the condition under which he works is pLo. Balance is attained if he attributes his own answers to the hypothetical person: pLx, oLx or p~Lx, o~Lx.

Moreover, the theory notes that when the circumstances are such that pLx, o~Lx or vice versa, balance results if p~Lo. Jordan has noted that all of his triadic relationships that have the component p~Lo are rated as being quite unpleasant. It is conceivable that such a resolution of imbalance is less likely to occur than one that reduces the relation of p and o to the level of indifference. Under either circumstance o's valence for p is less for dissimilar persons than for similar persons. The trend toward attributing identical answers to the other person should be diminished as a consequence.

The results of the test of the first hypothesis conform to this logical analysis. Greater similarity is attributed to persons perceived as being similar than to persons perceived as being dissimilar. This result holds both when the attributed similarity and dissimilarity are on high valence and low valence characteristics.

The second hypothesis is concerned with the question of whether or not the relationship between degree of perceived similarity and extent of projected similarity is a function of the valence of the material to be projected. This situation can be translated into terms of Heider's system as follows: In a particular p, o, x triad, individual o and characteristic x have specific valences for p. The extent of the force generated such that p perceives either oLx or o~Lx is a function of the valence of o (as demonstrated earlier), and the valence of x. The more crucial and acceptable (positive valence) x is for p, the more likely is p to associate or disassociate x and o. If o is not particularly crucial for p, the forces generated to establish balance are minimal.

The experimental results are consistent with this interpretation. Greater differentiation is made between similar and dissimilar persons on high valence items than on low valence items.

It should be noted that the employment of the Allport-Vernon scales restricts the treatment of results to relativistic terms. It is reasonable to believe that a person who received his highest score on the theoretical scale is positively oriented toward theoretical matters. However, a low score for the same person on the religious scale does not necessarily imply a rejection of religious issues and activities. It might involve rejection, indifference, or mild acceptance. The only legitimate statement concerning the theoretical and religious scales in this case would be that the former stands higher in positive valence than the latter.

SUMMARY

An attempt was made to investigate some of the implications of Heider's concept of balance. The subjects were 104 students enrolled in undergraduate courses. On the basis of their answers to the items on the Revised Allport-Vernon Scale of Values, partially completed test booklets were prepared incorporating different degrees of similarity to each subject's original answers. The subjects were required to examine the test booklets and complete them in the way the hypothetical persons had.

The results were consistent with the hypotheses which stated that greater similarity is projected on similar persons than on dissimilar persons for both high and low valence materials and that the differential is significantly greater on high valence material than on low valence material.

REFERENCES

1. HEIDER, F. Attitudes and cognitive organization. *J. Psychol.*, 1946, **21**, 107–112.
2. HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.
3. HOROWITZ, M. W., LYONS, J., & PERLMUTTER,

H. V. Induction of forces in discussion groups. *Hum. Relat.*, 1951, **4,** 57–76.

4. JORDAN, N. Behavioral forces that are a function of attitudes and of cognitive organization. *Hum. Relat.*, 1953, **6,** 273–287.

5. SMITH, A. J. Similarity of values and its relation to acceptance and the projection of similarity. *J. Psychol.*, 1957, **43,** 251–260.

Received August 28, 1957.

FUTURE TIME PERSPECTIVE, OPTIMISM, AND ACADEMIC ACHIEVEMENT[1]

JOHN E. TEAHAN

Connecticut State Hospital

Previous studies have indicated that the high academic achiever seems to be characterized by more optimistic attitudes and a greater concern for future goals (2, 3, 4). The setting up of goals implies a temporal orientation that is geared to the future, since a person's goals imply expectations and anticipations of future successes. High academic achievers might therefore be expected to show an essentially future oriented approach to life, characterized by a time perspective with broader scope than that of the less successful student. Findings of this nature have already emerged from a study by Teahan and Drews (7) in which it was found that intellectually gifted and high achieving junior high school students had a more extensive total time perspective than low achieving students of similar intellectual levels. In the following study, however, only future time perspective was considered. It was hypothesized that high achievers would have a more predominant and extensive future time perspective than low achievers and that a positive relationship would be found between optimism and extensiveness of future time perspective.

METHOD

The subjects (*S*s) used in this investigation were 60 seventh- and eighth-grade boys. Thirty *S*s (high achievers) had received academic grades in the upper quartile of their classes for the previous two years; the remaining 30 *S*s (low achievers) had been in the lower quartile for the same period. Groups were equated for age and for socioeconomic status as determined by rating the occupations of their fathers according to the method developed by Warner, Meeker, and Eells (8). Intellectual level was measured by the California Primary Mental Maturity Test which was administered in group form. Although no attempt was made to control for differences in intelligence between high achievers (mean IQ 108.3) and low achievers (mean IQ 95.5), the correlations obtained between intellectual level and the various measures of time perspective were found to approach zero in all cases (ranging from very low minus to very low plus correlations).

Three types of instruments were used to measure time perspective. The first of these techniques was originally developed by Eson (1) and consists of asking the *S* to record 25 things he had thought about or talked about in the past two weeks. *S* then rated these items according to whether, at the time he had thought or talked about them, they referred to something in the past, present, or future.

The second method, a story-completion technique, was first used by LeShan (5) for collecting data on time perspective. *S* was required to write a story starting with a partially completed statement or series of statements. The following sentences were used in the present investigation: (*a*) At three o'clock one bright sunny afternoon in May, two men were out walking near the edge of town . . . ; (*b*) Joe is having a cup of coffee in a restaurant. He's thinking of the time to come when . . . .

In the third part of the examination, *S*s were presented with three TAT cards. These included Card 1 (boy with a violin), Card 12B (boy sitting alone before a cabin) and Card 14 (silhouette of a person in a window). The cards were presented one at a time in this order, and *S*s were instructed to "Write a story about this picture."

When the entire testing session was ended, the examiner went over each story (both the story-completions and TATs) with the *S*, and asked the following questions: "How much time was involved in the action of this story—not in writing it but in the action described? How long would it have all taken if it had really happened?" In addition to this, each TAT story was rated by the examiner on a five-point scale according to the amount of optimism reflected in the outcome. Two other judges re-rated the TAT stories of 33 *S*s (16 high and 17 low achievers) in a reliability check of ratings on this variable. The obtained correlations were .85, .91, and .91.

RESULTS

Reliability of Future Time Perspective Measures

Of all instruments used, the story-completion technique was least reliable in terms of correlations between scores based on different stories. When Story 1 was compared to Story 2, the obtained cor-

---

[1] This paper is based on a doctoral dissertation submitted to the department of psychology, Michigan State University, in 1957. It was also presented at the meeting of the APA in New York, August, 1957. The writer wishes to express his appreciation to his chairman, A. I. Rabin, and to the other members of the committee who aided in the completion of the study.

TABLE 1

COMPARISON OF HIGH AND LOW ACHIEVERS WITH RESPECT TO AMOUNT OF FUTURE TIME PERSPECTIVE

Instrument	High Achievers Medians	Low Achievers Medians	Z	P
Future References	13.5	9.2	2.77	.003
Story-Completion #1	4 hours	1.5 hours	2.03	.021
Story-Completion #2	3 days	6 hours	1.32	.093
All Story-Completions	3.5 days	1.5 days	1.69	.045
TAT Story #1	1 day	1.5 hours	1.82	.034
TAT Story #2	2.5 days	2 hours	2.02	.022
TAT Story #3	12 hours	1 hour	1.82	.034
All TAT Stories	6 months	14 hours	2.21	.013

Note.—The Mann-Whitney nonparametric test was used because of the great variability in the time measures obtained which ranged from 5 minutes to 60 years. Because of the fact that the results were all in the predicted direction, and could not have been explained if they were different from that which was hypothesized, the probabilities listed are for a one-tailed test of statistical significance.

relation failed to reach a level of statistical significance. However, comparison of the TAT stories gave statistically significant correlations in all cases but one, when the scores obtained on any particular card by either high achievers, low achievers, or high and low achievers combined, were compared with their scores on any of the two remaining cards. There was also a high degree of internal consistency with respect to the future references obtained from recent conversations. An odd versus even comparison was made of the items listed by each *S* (only future references were considered), and the obtained correlations were significant well above the .01 level in all cases.

Future Time Perspective and Achievement

The recent conversations of high achievers included a significantly greater number of future references ($p < .01$) than those of low achievers. Although high achievers were found to have a more extensive future time perspective on both story-completions, only one of these (Story 1) gave statistically significant differences. When the results of both story-completions were combined, however, the differences were again significant by one-tailed test. The high achievers also demonstrated significantly more extensive future time perspective on all three stories given to the TAT cards.

Optimism and Future Time Perspective

A rank-difference correlation was computed between future time perspective and optimism on the TAT cards. The obtained coefficients of correlation were significant when high and low achievers were compared separately (coefficients of .47 and .40 respectively), or when they were studied as a total group (.50). Highly optimistic people tend to be more future oriented.

DISCUSSION

The results of Eson's (1) technique seems to indicate that high academic achievers are predominantly "anteverts" in so far as their recent thoughts and conversations are concerned, i.e., they tend to look mostly to the future. High achievers were also found to have a more extensive future time perspective than low achievers on most of the projective tests. These findings may mean that low achievers have a short-range or emergency type of future orientation, while high achievers have a long-range or foresightful type. Thus while continued and concentrated effort and attention to school subjects may be tedious and boring to the low achiever, these chores to the high achiever may represent meaningful subgoals on a time continuum in which the final end is eventual vocational success.

Students high in future extension also appeared to be more optimisitc. This agrees with Lewin's (6) theories concerning the relationship between self-confidence, hope, and a planning for a better future. The person without hope and confidence concentrates his attention on his immediate surrounds gaining whatever satisfactions he can in the present. In similar fashion, the student who has a history of poor grades might perceive the future as one in which only marginal success could be expected. Thus optimism, or the hope that positive changes can be expected in the future, would be intimately related to the extensiveness of a person's plans.

REFERENCES

1. ESON, M. E. An analysis of time perspective and five age levels. Unpublished doctoral dissertation, Univer. of Chicago, 1951.
2. GOUGH, H. G. Predicting success in graduate training. Inst. of Pers. Assessment and Res., Univer. of Calif., Berkeley, 1952. (Mimeo.)
3. GOUGH, H. G. Some theoretical problems in the construction of practical assessment devices for the early identification of high level talent. Dept. of Psychol. and Inst. of Pers. Assessment and Res., Univer. of Calif., Berkeley, 1953. (Mimeo.)
4. GOUGH, H. G. What determines the academic achievement of high school students. *J. educ. Res.*, 1953, **46**, 321–331.
5. LESHAN, L. L. Time orientation and social class. *J. abnorm. soc. Psychol.*, 1952, **47**, 589–592.
6. LEWIN, K. Time perspective and morale. In G. Watson (Ed.), *Civilian morale*. New York: Houghton Mifflin, 1942.
7. TEAHAN, J. E., & DREWS, ELIZABETH M. Time perspective and academic achievement. Paper read at Michigan Acad. of Arts and Sci., Detroit, March, 1957.
8. WARNER, W. L., MEEKER, MARCIA, & EELLS, K. *Social class in America*. Chicago: Science Research Assoc., 1949.

Received September 30, 1957.

HYSTERICS AND DYSTHYMICS AS CRITERION GROUPS IN THE MEASURE OF INTROVERSION–EXTRAVERSION: A REJOINDER TO EYSENCK'S REPLY[1]

JOHN J. SIGAL
Hampstead Child-Therapy Clinic, London

KOLMAN H. STAR
New South Wales University of Technology, Newcastle University College, N.S.W. Australia

AND CYRIL M. FRANKS
New Jersey Neuro-Psychiatric Institute

We hope to demonstrate that it was Eysenck (2), not we (3), who committed errors of logic and arrived at incorrect conclusions.

It was not Jung, but Eysenck who has contended that hysterics may be grouped with psychopaths and who coined the term dysthymia to include disorders not included under the term psychasthenia. Our findings, which Eysenck has accepted and supported by further (though inappropriate) data, demonstrated that hysterics may be grouped with dysthymics as justifiably as with pscyhopaths, but should probably not be grouped with either on I-E (introversion–extraversion). Contrary to Eysenck's statements, this finding is as relevant to his theory as it is to Jung's.

Eysenck seems to have made an error in logic. He has not recognized the difference between a necessary and a sufficient condition in logical argument. In order to demonstrate that hysterics are extraverted, it is a necessary condition that they be "toward the extraverted end of the scale" compared to dysthymics (assuming dysthymics to be introverted)—but it is not a sufficient condition. Elsewhere (1, p. 11 ff.) Eysenck, in quoting Jung, clearly demonstrated that Jung never intended to define an extravert as a less introverted introvert. To establish that hysterics are extraverted, it must also be shown that they are "toward the extraverted end of the scale" with respect to normals or *ambiverts* as Eysenck has called them, precisely as portrayed by him (1, p. 30).

From an experimental point of view, the relative positions of hysterics and dysthymics with respect to normals is important in determining the range of the I-E continuum over which one may consider as valid tests constructed, or conclusions about theories tested, using the former groups. Our results suggest that the use of these groups would limit the range to the introverted side of the continuum, that it would provide no information about the validity of tests or theories in the extraverted end of the continuum, and that tests constructed using hysterics and dysthymics as criterion groups might produce markedly skewed distributions when used in the normal population. These groups, used as a pair are, therefore, inadequate as criterion groups in the measure of I-E—if the E scale is a valid measure of I-E. The position of the normal group is essential to this conclusion. If we switched from one of Eysenck's deductions from Jung's theory to the other, it was because of logical and experimental necessity, not from want of logical consistency.

Since Eysenck has incorrectly anticipated our reply to his first point and has accepted the essential correctness of our results, his second point merits little discussion. Even allowing for the error in the logic of his first point, a significant difference between hysterics and psychopaths on I-E could hardly be considered to agree with prediction. It was because *our findings did not agree with prediction* that the hysteric–psychopath group (extraverted in theory but not in fact) was broken into its component parts, thereby producing smaller numbers. We might, indeed, have doubted the value of results based on so few cases were they not found to be in agreement with the results of other studies.

Eysenck suggested that we deliberately withheld relevant data, and that inclusion of these data would have altered our findings and the conclusions possible from these findings. We refer to his remarks about Jensen's data and the mean scores of the group used in standardizing the MPI.

Jensen selected his patients on the basis of extreme scores on I-E and obtained the diagnoses of these extreme scorers only. Our patients came from an entirely different hospital and were selected on the basis of diagnosis. The data from two such studies cannot be meaningfully combined. It was

---

[1] Editorial policy severely limits the space available to us. We are therefore only able to deal with some of the major issues raised by Eysenck, and even then in but a cursory manner.

Eysenck added his note of caution to ours (3) concerning the inadequacy of our sample for testing the merits of the N scale. Our summary did not repeat this note of caution contained in the text, so it is as well that Eysenck did. As the title of our article and the contents of Eysenck's reply indicated, however, the N scale results were neither the main nor an important issue in the study, so they will not be discussed in this rejoinder.

Eysenck who erred in pooling Jensen's data with those of Claridge, for example, which were obtained from unselected diagnosed patients. The reason for our use of Star's normal group rather than the standardization population was that the exact instructions and method of administering the test to the former group were known and could be followed, whereas there was no guarantee of the uniformity of these conditions in the testing of the latter group. In any case, the difference between the mean E scores of the two groups was too slight to warrant mentioning (12.63 compared to 12.45), since it would merely have confirmed a point for which we already had sufficient outside evidence.

In the paragraph before the summary, Eysenck indicated acceptance of our results, if not of our conclusions. If, as he suggested in this paragraph, there is a negative correlation between E and N for high N scorers and if, as he suggested, our diagnostic procedure resulted in the selection of less neurotic hysterics than are generally found in hospitals, then our selection was biased in favor of his theory, and presumably the result of a procedure which eliminated this bias would be even less favorable to his theory.

Finally, it should be pointed out that accepting the validity of the MPI as a measure of I-E would seem the least parsimonious of the alternatives open to Eysenck since, by so doing, he casts doubt on much of his earlier work in which hysterics were found to be extraverted.

SUMMARY

Eysenck has accepted our results despite their alleged premature publication, but has disputed our interpretation of them. The most prematurely published result, using his criterion of sample size, showed psychopaths to be extraverted, as predicted by his theory. Presumably, he did not intend this to be a contentious point. A consideration of the other arguments advanced by Eysenck, some of which were supported by irrelevant data, enables us to reaffirm our original statement, namely, either hysterics and dysthymics cannot be used as criterion groups in the measure of introversion–extraversion, or the E scale is not a good measure of introversion–extraversion, or both statements are true.

REFERENCES

1. EYSENCK, H. J. *The structure of human personality.* London: Methuen, 1953.
2. EYSENCK, H. J. Hysterics and dysthymics as criterion groups in the measure of introversion—extraversion: A reply. *J. abnorm. soc. Psychol.*, 1958, **57,** 250–252.
3. SIGAL, J. J., STAR, K. H., & FRANKS, C. M. Hysterics and dysthymics as criterion groups in the measure of introversion—extraversion. *J. abnorm. soc. Psychol.*, 1958, **57,** 143–148.

Received July 23, 1958.